STAT 330 notes: Mathematical Statistics

Johnew Zhang

December 3, 2012

Contents

1 Introduction

Mathematical statistics is a study of discrete datas collecting from practices.

Important Dates: Homework due date- Oct 3rd, Oct 31st, Nov 28th.

2 Review

2.1 Example 1.1

This example is given in the supplement course note on the course website. In order to study the frequency of number of fumbles in a game, a poisson model will be used. First of all, a random variable X for the number of fumbles in a game will be assigned as a poisson distribution, $X \sim POISSON(\lambda)$. A probability mass function could be defined for X, such that $P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!}$. Hence $E[X] = \lambda$.

For estimating the mean, we can do the following:

$$\hat{\lambda} = \frac{\sum_{i=1}^{110} X_i}{110} = \bar{X} \approx 2.55$$

Here the $\hat{\lambda}$ is a special estimator - MLE.

2.2 Likelihood Function

The probability of the observed data as a function of unknown parameter is the likelihood function. Now let's consider what is the probability of the observed table for example 1.1. The independence of number of fumbles among different teams is a hidden assumption in our example so the likelihood function will be easy to define

$$L(\lambda) = \frac{110!}{8! \cdots 1!}(\frac{\lambda^0 e^{-\lambda}}{0!})^8 \cdots (\sum_{x=8}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!})^0$$

Well, we need to test the significance of λ (it is 3 from the example). The following is a likelihood ratio test.

$$H_0 : \lambda = \lambda_0$$

$$-2 \log \frac{L(\lambda_0)}{L(\hat{\lambda}_{MLE})} \rightarrow_{d=1} \chi^2(1)$$

Next, we have yet another example about Weibull distribution. Weibull distribution measures the times to failure. Failure rate is proportional to a power of time ($\beta =$ the power $+ 1$)

$$f(x) = \frac{\beta}{\theta^\beta} x^{\beta-1} e^{-(\frac{x}{\theta})^\beta}, x > 0, \beta > 0, \theta > 0$$

The above function is the pdf for Weibull distribution. If $\beta < 1$, failure rate is decreasing with time; if $\beta = 1$, failure rate is constant; if $\beta > 1$, failure rate is increasing with time.

OK, let's go back to the example 2 for relief times in hours for 20 patients receiving a pain killer. Below is a likelihood function for θ and β.

$$L(\beta, \theta) = \prod_{i=1}^{20} [\frac{\beta}{\theta^\beta} x_i^{\beta-1} e^{-(\frac{x_i}{\theta})^\beta}] = [\frac{\beta}{\theta^\beta}]^{20} (\prod_{i=1}^{20} x_i)^{\beta-1} e^{-\sum_{i=1}^{20} (\frac{x_i}{\theta})^\beta}$$

Now we have the likelihood function and then it is easy to get the estimated β, θ by setting the partial derivative to zero.

3 Random Variable

Before starting, we will introduce couple of notations.

$$S : \text{sample space}$$
$$B = \{A_1, A_2, \cdots\}$$

1. $P(A) \geq 0, \forall A \in B$
2. $P(S) = 1$
3. $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
4. $P(\bar{A}) = 1 - P(A)$
5. $S = A \cup \bar{A}$
6. $1 = P(S) = P(A) + P(\bar{A})$

Here are some observations:

- $A \in S$ so $A \subseteq S$. In other words, A occurs if the outcomes of the random experiment is in the set.

Next, we define conditional probability as the following:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{if } P(B) > 0$$

Now if A and B are independent,

$$P(A \cap B) = P(A) \times P(B)$$

then $P(A|B) = P(A)$

Random Variable A random variable X is a function from a sample space S to the real number $\mathbb{R}$, that is

$$X : S \to \mathbb{R}$$

such that $F(X \le x)$ is defined for all $x \in \mathbb{R}$

Properties

- non-decreasing
- $\lim_{x\to\infty} F = 1$ and $\lim_{x\to-\infty} F = 0$.
- F is a right-continuous function, i.e. $lim_{x\to a^+} F(x) = F(a)$

Discrete Random Variable S is discrete then X is discrete.

Probability Mass Function $f(x) = P(X = x) = F(x) - \lim_{\xi\to 0^+} F(x - \xi), \forall x \in \mathbb{R}$ (non-zero and sum is equal to 1)

Here is an example for discrete random variables. The example is on L3 slides. Suppose we have a red balls and b blue balls.

1. The pmf of choosing x red balls in the first n selection with replacement

$$f(x) = P[X = x] = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$$

(Hyperbolic distribution)

2. The pmf of choosing x red balls in the first n selection without replacement

$$f(x) = P[X = x] = \binom{n}{x} \frac{a}{a+b}^{x} (\frac{b}{a+b})^{n-x}$$

(binomial distribution)

3. The pmf of number of blue balls selected before obtaining a red ball

$$P[X = x] = (1-p)^x p$$

(geometric distribution)

4. The pmf of number of blue balls selected before obtaining kth red ball

$$P[X = x] = \binom{x+k-1}{x}(1-p)^x p^k$$

(negative binomial distribution)

Example for Poisson Proof the sum of the p.m.f is 1.

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \frac{\mu^x e^{-\mu}}{x!}$$
$$= e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{-\mu} e^{\mu} = 1$$

(The proof is simple since when binomial distribution's n goes to infinity, it is poisson. in other words, $np = \mu$)

Continuous Random Variable S is continuous then X is continuous.

Probability Density Function The p.d.f has couple of properties: non-negative; sum of all is 1; sum to x we can get $F(x)$

Example 4 on L3 is an example for uniform distribution.

Gamma Function The gamma function denoted by $\Gamma(\alpha)$ for all $\alpha > 0$, is given by

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-\alpha} dy$$

1. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \alpha > 1$
2. $\Gamma(n) = (n - 1)!, n \in \mathbb{Z}_+$
3. $\Gamma(1/2) = \sqrt{\pi}$

The pdf of Gamma distribution is below

$$f(x) - \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}, x > 0, \alpha, \beta > 0$$

$X \sim GAM(\alpha, \beta)$. Special case, if $\alpha = 1$, then it is an exponential distribution.

3.1 Continuous Random Variable

If $Z \sim N(0, 1)$ find that the p.d.f. of $Y = Z^2$. If $X \sim N(\mu, \sigma^2)$ what is the distribution of $W = (\frac{X-\mu)^2}{\sigma})^2$?

$$G(y) = P(Y \leq y) = P(-\sqrt{y} \leq Z \leq \sqrt{y})$$
$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$
$$= 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Above is the c.d.f. for Y. To get p.d.f., we take the derivative.

$$g(y) = \frac{1}{2\sqrt{\pi y}} e^{-\frac{y}{2}} \sim \chi^2(1)$$

3.1.1 Probability Integral Transformation

If X is a continuous random variable with c.d.f. F then $Y = F(X) \sim UNIF(0,1)$. $Y = F(X)$ is called the probability integral transformation.

First of all, it is easy to get $f = 1$ and $F = x$.

Check the c.d.f for y

$$G(y) = P[Y \leq y] = P[F(X) \leq y] = P[X \leq F^{-1}(y)] = F(F^{-1}(y)) = y$$

3.1.2 Inversion Method

Suppose F is a c.d.f. for a continuous random variable. Show that if $U \sim UNIF(0,1)$ then the random variable $X = F^{-1}(U)$ also has c.d.f. F.

$$X = F^{-1}(U)$$
$$G(x) = P[X \leq x] = P[F^{-1}(U) \leq x] = P[U \leq F(x)] = F(x)$$

Above proves the statement. For simulation purpose, we wish to draw random number between 0 and 1. For each set we generated, we call it U_i. We can then convert all of those data to X_i data set which has the distribution of F.

3.1.3 One-to-One Transformation of a Random Variable

Suppose X is a continuous random variable with p.d.f. f and support set $A = \{x : f(x) > 0\}$ and $Y = h(X)$ where h is a real valued function. Let g be the p.d.f of the random variable Y and let $B = \{y : g(y) > 0\}$. If h is one-to-one function from A to B and h' is continuous, then

$$g(y) = f(h^{-1}(y)) \cdot |\frac{d}{dy} h^{-1}(y)|, y \in B$$

3.2 Expectation

If X is a discrete random variable with p.m.f. $f(x)$ and support set A then the expectation of X or the expected value of X is defined by

$$E[X] = \sum_{x \in A} x f(x)$$

provided the sum converges absolutely.

If X is a continuous random variable with p.d.f. $f(x)$ and support set A then the expected of X or the expected value of X is defined by

$$E[X] = \int_{x \in A} x f(x) dx$$

Example 1

$$X \sim GEO(p), 0 < p < 1$$
$$E(X) = \sum_A x f(x) = \sum_A x p (1-p)^x$$

Since $\sum_{x=0}^{\infty} x t^{x-1} = \frac{1}{1-t^2}, |t| < 1$,

$$E[X] = p(1-p) \sum_{x=0}^{\infty} x(1-p)^{x-1} = \frac{1-p}{p}$$

Theorem. If X is a discrete random variable with p.f $f(x)$ and support set A then

$$E[h(X)] = \sum_{x \in A} h(x) f(x)$$

provided the sum converges absolutely.

If X is a continuous random variable with p.d.f. $f(x)$ then

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) dx$$

provided the integral converges absolutely.

Theorem. Suppose X is a random variable with p.f/p.d.f. $f(x)$, a and b are real constants, and $g(x)$ and $h(x)$ are real -valued functions. Then

$$E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)]$$

Theorem.

$$Var(X) = E[X^2] - \mu^2 = E[X(X-1)] + \mu - \mu^2$$
$$Var(aX + b) = a^2 Var(X)$$

and

$$E[X^2] = \sigma^2 + \mu^2$$

Example 2

$X \sim BIN(n,p)$

$$\begin{aligned} E(X^{(k)} &= \sum_{x=k}^{\infty} h(x)f(x) & h(x) = X(X-1)\cdots(X-k+1) \\ &= \sum_{x=k} x^{(k)} \binom{n}{x} p^x (1-p)^{n-x} & x^{(k)}\binom{n}{x} = n^{(k)}\binom{n-k}{x-k} \\ &= \sum_{x=k} n^{(k)} \binom{n-k}{x-k} p^x (1-p)^{n-x} \\ &= \sum_{y=0} n^{(k)} \binom{n-k}{y} p^{y+k} (1-p)^{n-(y+k)} \\ &= n^{(k)} p^k \end{aligned}$$

3.2.1 Markov's Inequality

$$P(|X| \geq c) \leq \frac{E[(|X|^k]}{c^k}, \forall k, c > 0$$

Proof.

$$\begin{aligned} \frac{E[|X|^k]}{c^k} &= E[|\frac{X}{c}|^k|] \\ &= \int_{-\infty}^{\infty} |\frac{x}{c}|^k f(x)dx \\ &= \int_{\{|\frac{x}{c}| \geq 1\}} |\frac{x}{c}|^k f(x)dx + \int_{\{|\frac{x}{c}| < 1\}} |\frac{x}{c}|^k f(x)dx \\ &\geq \int_{\{|\frac{x}{c}| \geq 1\}} |\frac{x}{c}|^k f(x)dx \geq \int_{\{|\frac{x}{c}| \geq 1\}} 1 f(x)dx \\ &= P[|\frac{X}{c} \geq 1] = P[|X| \geq c] \\ P[|X| \geq c] &\leq \frac{E[|X|^k]}{c^k} \end{aligned}$$

□

3.2.2 Variance Transform

$$\begin{aligned}
X : E[X] = \theta,\, Var(X) = \sigma^2(\theta) \\
Y = g(X),\, Var(Y) \text{ is free of } \theta \\
Y = g(X) \approx g(\theta) + g'(\theta)(x - \theta) \\
E[Y] \approx g(\theta) + g'(\theta)E[X - \theta] = g(\theta) + g'(\theta)(E[X] - \theta) = g(\theta) \\
Var[Y] = [g'(\theta)]^2 Var(X - \theta) = [g'(\theta)]^2 \cdot Var(X)
\end{aligned}$$

If $Var[Y]$ is free of "θ", then

$$[g'(\theta)]^2 \cdot \sigma^2(\theta) \propto \text{constant}$$

Hence

$$[g'(\theta)]^2 \propto \frac{1}{\sigma^2(\theta)} \implies g'(\theta) \propto \frac{1}{\theta}$$

For example, $X \sim POIS(\theta), E[X] = \theta, Var[X] = \theta$. $Y = g(x), Var(Y) \propto$ constant (free of θ) $g'(\theta) \propto \frac{1}{\sqrt{\theta}}$
Actually, there are some nice properties for poisson random variable

1. $E[X^{(k)}] = \theta^k, k = 1, 2, 3, \cdots$

$$\begin{aligned}
E[X^{(k)}] &= \sum_{x=k}^{\infty} x^{(k)} \frac{e^{-\theta}\theta^x}{x!} \\
&= \sum_{x=k}^{\infty} \frac{e^{-\theta}\theta^x x(x-1)\cdots(x-k+1)}{x(x-1)x(-k+1)(x-k)\cdots 1} \\
&= \sum_{x=k}^{\infty} \frac{e^{-\theta}\theta^x}{(x-k)!} = \sum_{x=k}^{\infty} \frac{e^{-\theta}\theta^{x-k}}{(x-k)!}\theta^k \\
&= \theta^k \sum_{y=0}^{\infty} \frac{e^{-\theta}e^y}{y!} = \theta^k
\end{aligned}$$

3.3 Moment Generating Function

If X is a random variable then $M(t) = E[e^{tX}]$ is called the moment generating function (m.g.f.) of X if this expectation exists for all $t \in (-h, h)$ for some $h > 0$.

Example 1 $X \sim GAM(\alpha, \beta)$ then find $M(t)$

$$
\begin{aligned}
M(t) = E[e^{tX}] &= \int_0^\infty e^{tx} f(x)dx \\
&= \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx \\
&= \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x(1/\beta - t)} dx \\
&= \int \frac{1}{\beta^\alpha \Gamma(\alpha)} (y \frac{\beta}{1-\beta t})^{\alpha-1} e^{-y} (\frac{\beta}{1-\beta t}) dy \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \frac{\beta^\alpha}{(1-\beta t)^\alpha} \int y^{\alpha-1} e^{-y} dy
\end{aligned}
$$

When $\frac{1}{\beta} > t$,

$$M(t) = \frac{1}{(1-\beta t)^\alpha}$$

Example 2 $X \sim NB(k, p)$, then $M(t)$

$$
\begin{aligned}
M(t) = E(e^{tX}] &= \sum_{k=0}^{\infty} e^{tx} \binom{k+x-1}{x} p^k (1-p)^x \\
&(\binom{k+x-1}{x} = (-1)^x \binom{-k}{x}) \\
&= \sum_{k=0}^{M} \infty \binom{-k}{x} p^k (e^t(p-1))^x \\
&= p^k \sum_{k=0}^{\infty} \binom{-k}{x} (e^t(p-1))^x \\
&= p^k (1 + e^t(p-1)^{-k} \text{ if } |e^t(p-1)| < 1
\end{aligned}
$$

Theorem. Suppose the random variable X has m.g.f. $M_X(t)$ defined for $t \in (-h, h)$ for some $h > 0$. Let $Y = aX + b$ where $a, b \in \mathbb{R}$ and $a \neq 0$. Then the m.g.f of Y is

$$M_X(t) = e^{bt} M_X(\alpha t), |t| < \frac{b}{|a|}$$

Proof.

$$M_Y(t) = E[e^{tY}] = E[e^{t(aX+b)}] = E[e^{tax+tb}] = e^{tb} E[e^{atx}]$$

□

Example 3 $X \sim N(\mu, \sigma^2)$, then $M(t)$.

Since $Z = \frac{x-\mu}{\sigma} \sim N(0,1)$, then $X = \sigma Z + \mu$.

Step 1. Find $M_Z(t), h$

Step 2. $X = \sigma Z + \mu$.

Step 3. Apply theorem above $M_X(t) = e^{\mu t} M_Z(\sigma t), |t| < \frac{h}{a}$

Theorem. Suppose the random variable X has m.g.f. $M(t)$ defined for $t \in (-h, h)$ for some $h > 0$. Then $M(0) = 1$ and

$$M^{(k)}(0) = E(X^k), k = 1, 2, 3, \cdots$$

Theorem. Suppose the random variable X has m.g.f. $M_X(t)$ and the random variable Y has m.g.f. $M_Y(t)$. Suppose also that $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$. Then X and Y have the same distribution, that is $P(X \leq s) = F_X(s) = F_Y(s) = P(Y \leq s)$ for all $s \in \mathbb{R}$

3.4 Joint Probability

Example 1

Suppose we have 10 acts students, 9 stat students and 6 math business students. We select 5 out of 25 without replacement. Suppose $X =$ No. ACTSC, $Y =$ No. STAT. Suppose $A = \{X > Y\}$

1. Joint p.m.f of X and Y.

$$P[X = x, Y = y] = \frac{\binom{10}{x}\binom{9}{y}\binom{6}{5-x-y}}{\binom{25}{5}}$$

2.

$$P[X = x] = \sum_{y=0}^{5-x} P[X = x, Y = y] = \frac{\binom{10}{x}\binom{15}{5-x}}{\binom{25}{5}}$$

3.

$$P[Y = y] = \sum_{x=0}^{5-y} P[X = x, Y = y] = \frac{\binom{9}{y}\binom{16}{5-y}}{\binom{25}{5}}$$

Example 2

Suppose X and Y are continuous random variable with joint p.d.f

$$f(x, y) = x + y, 0 \leq x \leq 1, 0 \leq y \leq 1$$

and 0 otherwise. Show that the summation is 1.

$$\int_0^1 \int_0^1 x + y dy dx = \int_0^1 xy + \frac{y^2}{2}|_0^1 dx = \int_0^1 x + \frac{1}{2} dx = \frac{x^2}{2} + \frac{1}{2}x|_0^1 = 1$$

Calculate $P(X \leq Y)$

$$P(X \leq Y) = \int_0^1 \int_0^y f(x, y) dx dy = 0.5$$

Calculate $P(X + Y \leq \frac{1}{2})$

$$P(X + Y \leq \frac{1}{2}) = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2} - y} f(x, y) dx dy = \frac{1}{24}$$

Calculate $P(XY \leq \frac{1}{2})$

$$P(XY \leq \frac{1}{2}) = 1 - \int_{\frac{1}{2}}^1 \int_{\frac{1}{2x}}^1 f(xy) dx dy = \frac{3}{4}$$

Find joint c.d.f. of (X, Y)

$$F(xy) = \int_0^x \int_0^y f(s, t) ds dt$$

Independent Random Variables

Two random variables X and Y are called independent random variables if

$$P(X \in A \text{ and } Y \in B) = P(X \in A) P(Y \in B)$$

for all sets A and B of real numbers.

Theorem. Suppose X and Y are random variables with joint c.d.f $F(x, y)$ and joint p.d.f. $f(x, y)$, marginal c.d.f's $F_X(x)$ and $F_Y(y)$ respectively, and marginal p.d.f $f_X(x)$ and $f_Y(y)$ respectively. Suppose that $A_1 = \{x : h(x) > 0\}$ is the support set of X and $A_2 = \{y : f_Y(y) > 0$ is the support set of Y.

Example

The student selection example:

$$f(x,y)=\frac{\binom{10}{x}\binom{9}{y}\binom{6}{5-x-y}}{\binom{25}{5}}, x=0,1,2,\cdots,5, y=0,1,\cdots,5, x+y\le 5$$

$$f_1(x)=\frac{\binom{10}{x}\binom{15}{5-x}}{\binom{25}{5}}$$

$$f_2(y)=\frac{\binom{9}{y}\binom{16}{5-y}}{\binom{25}{5}}$$

Easy to verify that x and y are not independent.

3.5 Factorization Theorem for Independence

Example

Suppose X and Y are discrete random variable with p.m.f

$$f(x,y)=\frac{\theta^{x+y}e^{-2\theta}}{x!y!}, x=0,1,\cdots,y=0,1,\cdots$$

Are X and Y independent random variables? Find the marginal p.m.f. of X and the marginal p.m.f.

$A=\{(x,y):x=0,1,\cdots,y=0,\cdots\}$, $A_1=\{x:x=0,1,\cdots\}$, $A_2=\{y:y=0,1,\cdots\}$. Well, $A=A_1\times A_2$.

$$f(x,y)=g(x)h(y)=\frac{\theta^x e^{-\theta}}{x!}\frac{\theta^y e^{-\theta}}{y!}$$

Hence it is independent.

Example

Suppose X and Y are continuous random variables with joint p.d.f

$$f(x,y)=\frac{2}{\pi}, 0\le x\le\sqrt{1-y^2}, -1\le y\le 1$$

and 0 otherwise. Are X and Y independent random variables? Find the marginal p.d.f of X and the marginal p.d.f of Y.

$$f_1(x)=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x,y)dy=2\int_0^{\sqrt{1-x^2}}\frac{\pi}{2}dy=\frac{4}{\pi}\sqrt{1-x^2}, 0<x<1$$

$$f_2(x)=\int_0^{\sqrt{1-y^2}} f(xy)dx=\frac{\pi}{2}\sqrt{1-y^2}, -1<y<1$$

Proof. Proof for the Factorization theorem: From LHS, if X and Y are independent, $f(xy) = f(x)f(y)$. Let $g(x) = f_1(x)$, $h(y) = f_2(y)$. From RHS, Let $c = \int_{A_1} g(x)dx, d = \int_{A_2} f(y)dy, c, d > 0, f_1(x) = \int_{A_2} f(x,y)dy = \int_{A_2} g(x)h(y)dy = g(x)d$

$f_2(y) = eh(y)$, $\int_{A_2}\int_{A_1} f(x,y)dxdy = 1$, $\int_{A_2}\int_{A_1} g(x)h(y)dxdy = cd$. □

3.6 Conditional Joint Probability

Example

$$\begin{cases} f(xy) = \frac{2}{\pi}, & 0 \le x \le \sqrt{1-y^2}, -1 < y < 1 \\ f_1(x) = \frac{4}{\pi}\sqrt{1-x^2} & 0 \le x \le 1 \\ f_2(y) = \frac{2}{\pi}\sqrt{1-y^2} & -1 < x < 1 \end{cases}$$

$\implies f_1(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{1}{\sqrt{1-y^2}}, 0 \le x \le \sqrt{1-y^2}, -1 < y < 1$

Product rule: $f(x,y) = f_1(x|y)f_2(y) = f_2(y|x)f_1(x)$

Example

$X|Y = y \sim BIN(y,p), Y \sim POI(u)$. Marginal pmf of X

$f_1(x|y) = \binom{y}{x}p^x(1-p)^{y-x}$ and $f_2(y) = \frac{e^{-u}u^y}{y!}$.

$f(x,y) = \binom{y}{x}p^x(1-p)^{y-x}\frac{e^{-u}u^{-y}}{y!}$

$f_1(x) = \sum_{y=x}^{\infty} f(x,y)$

3.7 Conditional Expectation

Example

$f(x,y) = \frac{2}{\pi}, 0 < x < \sqrt{1-y^2}, -1 < y < 1$. Find $E[Y|X], Var(Y|X)$.

$$f_2(y|x) = \frac{1}{2\sqrt{1-x^2}}, 0 < x < 1, -\sqrt{1-x^2} < y < \sqrt{1-x^2}$$

$E[Y|X] = \int yf_2(y|x)dy = 0$.

$E[Y^2|X] = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2 f_2(y|x)dx = \frac{1}{3}(1-x^2)$

$Var(Y|X) = E[Y^2|X] - E[Y|X]^2$

$$E[E[X|Y]] = E[X]$$

$$Var[Y] = Var[E[Y|X]] + E[Var[Y|X]]$$

Prove above yourself (it is really easy)

3.8 Joint Moment Generating Function for X and Y

$$M(t_1, t_2) = E[e^{t_1 X + t_2 Y}], \text{ for } t_1 \in (-h_1, h_1), (-h_2, h_2)$$

Joint mgf for $x_1, x_2, \cdots, x_n 4 some \text{h}_1, h_2 > 0$

$$M(t_1, t_2, \cdots, t_n) = E[e^{t_1 X_1 + \cdots t_n X_n}], \text{ for } t_i \in (-h_i, h_i), h_i > 0$$

$$M(t, 0) = E[e^{t_1 X + 0Y}] = E[e^{t_1 X}] = M_X(t_1), t_1 \in (-h_1, h_1)$$

$$M(0, t_2) = M_Y(t_2), t_2 \in (-h_2, h_2)$$

Recall: $E[X^k] = M_X^{(k)}(0)$ Therefore

$$E[X^k Y^j] = \frac{\partial^{k+j}}{\partial X^k \partial Y^j} M(t_1, t_2)|_{(t_1, t_2)=(0,0)}$$

3.8.1 The independence Theorem of Moment Generating Function

X and Y are independent if and only

$$M(t_1, t_2) = M_X(t_1) M_Y(t_2)$$

Example: $f(x, y) = e^{-y}, 0 < x < y < \infty$. Find joint mgf of X and Y.

$$\begin{aligned}
M(t_1, t_2) = E[e^{t_1 X + t_2 Y}] &= \int_0^\infty \int_0^y e^{t_1 x + t_2 y} f(x, y) dx dy \\
&= \int_0^\infty \int_0^y e^{t_1 x + t_2 y} e^{-y} dx dy = \int_0^\infty e^{(t_2 - 1)y} \int_0^y e^{t_1 x} dx dy \\
&= \int_0^\infty e^{(t_2 - 1)y} [\frac{1}{t_1} e^{t_1 x}|_0^y] dy = \frac{1}{t_1} \int_0^\infty e^{(t_2 - 1)y} [e^{t_1 y} - 1] dy \\
&= \frac{1}{t_1} (\frac{1}{t_2 - 1} - \frac{1}{t_1 + t_2 - 1}), \text{ if } (t_2 - 1) < 0, \text{ and } (t_1 + t_2 - 1) < 0 \\
&= \frac{1}{(1 - t_2)(1 - t_1 - t_2)}
\end{aligned}$$

$$M_X(t) = M(t, 0) = \frac{1}{1 - t}, t < 1$$

$$M_Y(t) = M(0, t) = \frac{1}{(1 - t)^2}, t < 1$$

Therefore $X \sim Exp(1), Y \sim Gamma(2, 1)$. Well $M(t_1, t_2) \neq M_X(t_1) M_Y(t_2)$ for all $t_2 < 1$ and $t_1 + t_2 < 1$ so X and Y are not independent.

3.9 Multinomial Random Variables $(X_1, \cdots, X_k) \sim Multinomial(n, P_1, P_2, \cdots P_k)$

$$f(x_1, \cdots, x_k, x_{k+1}) = \frac{n!}{x_1!x_2!\cdots x_k!x_{k+1}!}P_1^{x_1}\cdots P_{k+1}^{x_{k+1}}$$

where $x_{k+1} = n - (x_1 + \cdots + x_k), P_{k+1} = 1 - (P_1 + P_2 + \cdots + P_k)$

1. $M(t_1, \cdots t_k) = E[e^{t_1X_1+\cdots t_kX_k}]$, For the case $k = 2$, $M(t_1, t_2) = E[e^{t_1X_1+t_2X_2}] = \sum\sum e^{t_1x_1+t_2x_2}\frac{n!}{x_1!x_2!(n-x_1-x_2)!}$

2. $Cov(x_i, x_j) = -np_ip_j$.

3.10 Bivariate Normal Distribution

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim BVN(\mu, \Sigma) = BVN(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho(\sigma_1, \sigma_2) \\ \rho(\sigma_1, \sigma_2) & \sigma_2^2 \end{pmatrix})$$

$$f(x_1, x_2) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

Properties:

1. $c^TX \sim N(c^T\mu, c^T\Sigma c)$ for constant vector $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ where $c^T\mu = c_1\mu_1 + c_2\mu_2$.

   $M_Y(t) = E[e^{tY}] = E[e^{t(c^Tx)}] = E[e^{(ct)^Tx}]$ where t is a scaler.

Recall If $Z \sim N(\mu, \sigma^2)$, $M_Z(t) = EXP(\mu + \frac{1}{\sigma}t^2\sigma^2)$. Due to the uniqueness of mgr, then we know $Y \sim N(c^Tu, c^T\Sigma c)$.

4 Function of Random Variables

4.1 CDF technique

Example

$f(x, y) = 3y, 0 \leq x \leq y \leq 1$ and $T = X \cdot Y$. Using the cdf technique:

If $t < 0$, $G(t) = P[T \leq t] = 0$. If $t \geq 1$, $G(t) = P[T \leq t] = 1$. For $t \in (0, 1)$, $G(t) = \int\int_{\mathbb{R}} f(x, y)dxdy = 3t - 2t\sqrt{t}$. Therefore $g(t) = G'(t) = 3(1 - \sqrt{t}), t \in (0, 1)$.

Example

$X_1, \cdots, X_n \sim F(x)$ iid.

$$\begin{cases} X_{(n)} = \max\{X_1, \cdots, X_n\} \\ X_{(1)} = \min\{X_1, \cdots, X_n\} \end{cases}$$

pdf's of $X_{(n)}$ and $X_{(1)}$?

$$\begin{cases} \text{if } X_{(n)} \le y & \iff x_i \le y \text{for all } i = 1, \cdots, n \\ \text{if } X_{(1)} > y & \iff x_i > y \text{ for all } i = 1, \cdots n \end{cases}$$

$$\begin{aligned} G(y) &= P[X_{(n)} \le y] = P(X_1 \le y, X_2 \le y, \cdots, X_n \le y] \\ &= \prod_{i=1}^{n} P[A_i] = \prod_{i=1}^{n} P[X_i \le y] = \prod_{i=1}^{n} [F(y)] = [F(y)]^n \end{aligned}$$

Hence

$$g(y) = \frac{d}{dy} G(y) = n[F(y)]^{n-1} F(y)' = n[F(y)]^{n-1} f(y)$$

where $g(y)$ is the pdf of $X_{(n)}$.

On the other hand, for $X_{(1)}$

$$\begin{aligned} H(y) &= P[X_{(1)} \le y] = 1 - P[X_{(1)} > y] = 1 - P(X_1 > y, X_2 > y, \cdots, X_n > y] \\ &= 1 - \prod_{i=1}^{n} P(B_i) = 1 - \prod_{i=1}^{n} P[X_i > y] = 1 - (1 - F(y))^n \end{aligned}$$

Therefore, $h(y) = n(1 - F(y))^{n-1} f(y)$

4.2 One-to-One transformation

$u = h_1(x, y), v = h_2(x, y)$ is one-to-one transformation form. Map $R_{XY} R_{uv}$, R_{XY} is the support set of $f(x, y)$. The inverse: $X = \omega_1(u, v), y = \omega_2(u, v)$.

One-to-One? Check

1. $\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}$ are continuous functions.

2. $\frac{\partial(u,v)}{\partial(x,y)} = det[\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}] \ne 0$ for all $(x, y) \in \mathbb{R}_{XY}$. pdf of (u, v), $g(u, v)$ is $g(u, v) = f(\omega_1(u, v), \omega_2(u, v)) |\frac{\partial(x,y)}{\partial(u,v)}|$

Example

$$X \sim Gamma(a, 1), f_1(x) = \frac{x^{a-1}}{\Gamma(a)} \frac{x^{a-1}}{\Gamma(a)} e^{-x}, x > 0$$

$$Y \sim Gamma(b, 1), f_2(x) = \frac{y^{b-1}}{\Gamma(b)} \frac{y^{b-1}}{\Gamma(b)} e^{-y}, y > 0$$

Transformation: $U = X + Y$ and $V = \frac{X}{X+Y}$. Find the joint pdf of (U, V).

Check one-to-one transformation:

1. $\frac{\partial U}{\partial x} = 1$, $\frac{\partial U}{\partial y} = 1$, $\frac{\partial V}{\partial x} = \frac{Y}{(X+Y)^2}$, $\frac{\partial V}{\partial y} = \frac{-X}{(X+Y)^2}$
2. $\cdots$

$g(u,v) = f(w_1(u,v), w_2(u,v))|\frac{\partial(x,y)}{\partial(u,v)}, (u,v) \in R_{uv}$

1. $f(x,y) = f(x)f(y), x > 0, y > 0$ due to independence
2. $X = w_1(u,v) = uv$, $Y = w_2(u,v) = u - uv$
3. $\frac{\partial(x,y)}{\partial(u,v)} = -u$, then $|\cdots|$.

Therefore, $g(u,v) = f_1(w_1(u,v))f_2(w_2(u,v))u$ where $f_1(w_1(u,v))$ and $f_2(w_2(u,v)$ are the transformation of X and Y. Hence we get

$$g(u,v) = [\frac{u^{(a+b)-1}e^{-u}}{\Gamma(a+b)}]([\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}]v^{a-1}(1-v)^{b-1})$$

where the first half is a $Gamma(a+b)$ and the second half is a $Beta(a,b)$. Due to Factorization then U and V are independent and $R_{uv} = \{(u,v) : u > 0, 0 < v < 1\}$.

To find $E[U^2V^3] = E[U^2]E[V^3]$

Theorem. $Z \sim N(0,1)$, $X \sim \chi^2(n)$ and $\frac{Z}{\sqrt{\frac{X}{n}}} \sim t_{(n)}$

$$(Z, X) \iff (S = \frac{Z}{\sqrt{\frac{X}{n}}}, T = X)$$

Theorem. $X \sim \chi^2(n)$ and $Y \sim \chi^2(m)$. Hence $U = \frac{X/n}{Y/m} \sim F(n,m)$

4.3

$X_1, \cdots, X_n$ are $M_i(t)$ is the mgf of X_i, for some $h > 0$. $Y = \sum_{i=1}^n X_i$. Therefore $M_Y(t) = \prod_{i=1}^n M_i(t)$. If $X_1, \cdots X_n$ are i.i.d., then $[M(t)]^n$

For example in 4.2, $g(u,v) = f(w_1(u,v), w_2(u,v))|\frac{\partial(x,y)}{\partial(u,v)} =$ "1". Hence $f(x,y) = 1 \times 1$.

If $X_1, \cdots, X_n$ are independent. Let $M_i(t)$ be the mgr of X_i, $t \in (-h,h)$. Let $Y = \sum_{i=1}^n X_i$, $M_Y(t) = \prod_{i=1}^n M_i(t), t \in (-h,h)$. If $X_1, \cdots, X_n$ are i.i.d.,$M_Y(t) = [M(t)]^n, t \in (-h,h)$.

Special Results:

1. If $X \sim Gamma(\alpha, \beta), \alpha, \beta > 0, x > 0$, then $\frac{2X}{\beta} \sim \chi^2(2\alpha)$. $M_x(t) = \frac{1}{(1-\beta t)^2}, t < \frac{1}{\beta}$.
   $Y = \frac{2X}{\beta}, M_Y(t) = M_X(\frac{2t}{\beta}) = \frac{1}{(1-2t)^2}$
   Hence $\frac{1}{(1-2t)^2}$ is the mgr of $\chi^2(2\alpha)$ r.v. $t < \frac{1}{2}$.

2. $X_i \sim Gamma(\alpha_i, \beta), i = 1, \cdots, n$ independent.
   $\sum_{i=1}^n X_i \sim Gamma(\sum_{i=1}^n \alpha_i, \beta)$.
   $Y = \sum_{i=1}^n X_i$, $M_Y(t) = \prod_{i=1}^n M_i(t), M_i(t)$ is mgf of X_i.

$$M_i(t) = \frac{1}{(1-\beta t)^2}, M_Y(t) = \prod_{i=1}^n (\frac{1}{(1-\beta t)^2} = \frac{1}{(1-\beta t) \sum_{i=1}^n \alpha_i}$$

   Hence we get the mgf.

3. $X_i \sim Bin(n_i, p), i = 1, \cdots, n$ are independent.

$$\sum_{i=1}^n X_i \sim BIN(\sum_{i=1}^n n_i, p)$$

$$M_i(t) = (pe^t + (1-p))^{n_i}, t \in \mathbb{R}$$

$$Y = \sum_{i=1} X_i$$

$$M_Y(t) = \prod_{i=1}^n M_i(t) = (pe^t + (1-p))^{\sum n_i}$$

   is the mgf of $Bin(\sum_{i=1}^n n_i, p)$

Theorem. $X_i \sim N(\mu_i, \sigma_i^2)$ are independent.

$$\sum_{i=1}^n a_i X_i \sim N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$$

$X_i \sim (\mu, \sigma_i^2)$, $M_i(t) = Exp(t\mu^2 + \frac{1}{2} t^2 \sigma_i^2)$
Well, $Y = \sum a_i X_i$, so

$$M_Y(t) = exp(t \sum_{i=1}^n a_i + \frac{1}{2} t^2 (\sum a_i^2 \sigma_i^2))$$

is the mgf.

Corollary. $X_1, \cdots, X_n$ are i.i.d. $N(\mu, \sigma^2)$.

Then $\sum X_i \sim N(\sum a_i\mu_i, \sum a_i^2\sigma_i^2)$ where $a_i = 1$ for $i = 1, \cdots n$ and $\mu_1 = \mu_2 = \cdots = \mu_n = \mu$ and $\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2$

$$\bar{X} = \frac{\sum X_i}{n}$$

$$M_{\sum X_i}(t) = exp(t(n\mu) + \frac{1}{2}t^2(n\sigma^2))$$

$$M_{\bar{X}}(t) = exp(t\mu + \frac{1}{2}t^2(\frac{\sigma^2}{n}))$$

this is the mgf.

Theorem. $X_1, \cdots, X_n$ are i.i.d. $N(\mu, \sigma^2)$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ and $S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$ are independent. Hence

$$M(t_1, t_2) = M_{\bar{X}}(t_1)M_{S^2}(t_2)$$

Therefore

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Proof. Recall $\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$ and $[\frac{n(\bar{x} - \mu)^2}{\sigma^2} \sim \chi^2(1)$. Hence $M_{\chi^2(n)}(t) = M_{\frac{(n-1)s^2}{\sigma^2}}(t)M_{\chi^2(1)}(t)$

$$M_{\frac{(n-1)s^2}{\sigma^2}}(t) = \frac{M_{\chi^2(n)}(t)}{M_{\chi^2(1)}(t)} = \frac{\frac{1}{(1-2t)^{\frac{n}{2}}}}{\frac{1}{(1-2t)^{1/2}}}$$
$$= \frac{1}{(1-2t)^{\frac{n-1}{2}}}$$

is the mgf of $\chi^2(n-1)$

$\therefore \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$. □

Theorem. $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ iid

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

Proof. Recall $\frac{N(0,1)}{\sqrt{\frac{\chi^2(n)}{n}}} \sim t(n)$ and $Z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$ and $Y = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$.

$$\frac{Z}{\sqrt{Y/(n-1)}} \sim t(n-1)$$

Plug in the Z and Y we can easily derive $\frac{\bar{x} - \mu}{s/\sqrt{n}}$.

□

Theorem. $X_1, \cdots, X_n \sim N(\mu_1, \sigma_1^2)$ independent. $Y_1, \cdots, Y_m \sim N(\mu_1, \sigma_2^2)$ independent.

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F(n-1, m-1)$$

.

Proof. Recall: $X \sim \chi^2(n), Y \sim \chi^2(m)$ are independent and $\frac{X/n}{Y/m} \sim F(n, m)$.

$\frac{(n-1)s_1^2}{\sigma_1^2} \sim \chi^2(n-1)$ and $\frac{(n-1)s_2^2}{\sigma_2^2} \sim \chi^2(m-1)$.

Therefore $s_1^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$ and $s_2^2 = \sum_{i=1}^m \frac{(y_i - \bar{y})^2}{m-1}$

Therefore $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F(n-1, m-1)$. □

5 Convergence

5.1 Convergence in distribution

5.1.1 e-limit

$$\lim_{n\to\infty} (1 + \frac{k}{n})^{cn} = e^{bc}$$

b and c are constants!

$$\lim_{n\to\infty} F_n(x) = F(x)$$

, for all x at which $F(\cdot)$ is continuous.

$$\lim_{n\to\infty} (1 - \frac{2}{n})^n = \lim_{n\to\infty} (1 + \frac{(-2)}{n})^n = e^{-2}$$

If at $x = x_0$, $\lim_{n\to\infty} F_n(x_0) \neq 0$, but $\lim_{n\to\infty} F_n(x) = F(x)$, for x at which $F(\cdot)$ is continuous.

$$X_n \to_d X \sim F(x)$$

Example: $X_1, \cdots, X_n, \cdots \sim Exp(1)$ iid. $Y_n = \max(X_1, \cdots, X_n\} - \log(n)$

Find the limiting distribution U_n

1. cdf of $Y_n, F_n(y)$
2. $\lim_{n\to\infty} F_n(y) =?$

The pdf of X_i is $e^{-x}, x > 0$, $\therefore$ the cdf of X_i is $\int_0^x e^{-t} dt = 1 - e^{-x}, x > 0$. The cdf of X_i is 0 if $x \leq 0$. The cdf of $Y_n, F_n(y)$

$$\begin{aligned}
F_n(y) = P[Y_n \leq y] &= P[\max\{X_1, \cdots, X_n\} - \log(n) \leq y\}] = P[\max\{X_1, \cdots, X_n\} \leq y + \log n] \\
&= P[X_1 \leq y + \log(n), \cdots, X_n \leq y + \log(n)] \\
&= \prod_{i=1}^n P[X_i \leq y + \log(n)] = \prod_{i=1}^n F_i[y + \log n]
\end{aligned}$$

If $y + \log n > 0$,

$$\begin{aligned}
F_n(y) \neq 0 = \prod_{i=1}^{n} F_i(y + \log n) &= \prod_{i=1}^{n}(1 - e^{-(y+\log(n))}) \\
&= (1 - e^{-(y+\log(n))})^n \\
\lim_{n\to\infty} F_n(y) = \lim_{n\to\infty}(1 - e^{-(y+\log(n))})^n &= \lim_{n\to\infty}(1 - e^{-y}/n)^n \\
&= e^{-e^{-y}} = F(y)
\end{aligned}$$

If $Y \sim F(y), Y_n \to^d Y$

5.2 Convergence in Probability

$X_1, X_2, \cdots, X_n, \cdots$. $\lim P(|X_n - X| < \epsilon) = 1$ or $\lim P[|X_n - X| \geq \epsilon] = 0$. (

Theorem. If $X_n \to^P X$, then $X_n \to_d X$

)
$X_n \to^P X$, $X_n \to^P b$ is constant!.

$$\lim_{n\to\infty} P[|X_n - b| < \epsilon] = 1$$

Theorem. If $X_n \to^d b$, constant! then $X_n \to^P b$.

Now if $X_n \to^d b$, constant! $\lim_{n\to\infty} F_n(x) = F(x)$, where $F(x)$ is the cdf of b.

$$F(x) = \begin{cases} 0 & \text{if } x < b \\ 1 & \text{if } x = b \\ 1 & \text{if } x > b \end{cases}$$

5.2.1 Example

$X_i \sim Exp(1, 0), i = 1, 2, \cdots$ independent. $Y_n = \min\{X_1, \cdots, X_n\}$ and $Y_n \to^P \theta$.

$$\begin{cases} Y_1 = X_1 \\ Y_2 = \min\{X_1, X_2\} \\ Y_n = \min\{X_1, \cdots X_n\} \end{cases}$$

Y's might be dependent.

$$F_n(y) = P[Y_n \leq y] = 1 - P[Y_n > y] = 1 - P[X_1 > y, X_2 > y, \cdots, X_n > y] = 1 - \prod_{i=1}^{n} P[X_i > y]$$

$X_i \sim Exp(1,\theta), i = 1,2,\cdots$ and the pdf $h_i(x) = e^{-(x-\theta),x>\theta}$ and the cdf $H_i(x) = \int_\theta^x h_i(t)dt = 1 - e^{-(x-\theta)}$ for $x > 0$. If $x \leq 0$, $H_i(x) = 0$.

$$F_n(y) = 1 - \prod_{i=1}^{n}(1 - H_i(y)) = 1 - (1 - H_1(y))^n = 1 - (1 - (1 - e^{-(y-\theta)}))^n = 1 - e^{-n(y-\theta)}$$

$$\lim_{n\to\infty} F_n(y) = \lim_{n\to\infty}(1 - e^{-n(y-\theta)}) = 1 \text{ for } y > \theta$$

$$F(y) = \begin{cases} 0 & \text{if } y < \theta \\ 1 & \text{if } y > \theta \end{cases}$$

Hence $Y_n \to^d Y = \theta$; therefore $Y_n \to^P Y = \theta$

5.3 Limit Theorem

$X_n \to^d X$ and $\lim_{n\to\infty} F_n(x) = F(x)$, Well let X_n's mgf be $M_n(t)$ and X's mgf be $M(t)$. Hence

$$\lim_{n\to\infty} M_n(t) = M(t), t \in (-h,h)$$

5.3.1 Example

$X_n \sim BIN(n,p)$ and $M_n(t) = (pe^t + 1 - p)^n, t \in \mathbb{R}$, What is the $\lim_{n\to\infty} M_n(t) =$? $np = \mu, n \to \infty, p \to 0$.

$$\lim_{n\to\infty} M_n(t) = \lim_{n\to\infty}(\frac{\mu}{n}e^t + 1 - \frac{\mu}{n})^n = \lim_{n\to\infty}(1 + \frac{\mu e^t - 1}{n})^n = e^{\mu(e^t-1)}$$

Due to the uniqueness of moment generating function, $X_n \to^d X \sim Poi(\mu)$

5.4 Central Limit Theorem

X_i are i.i.d. r.vs. $E[X_i] = \mu < \infty$ and $Var(X_i) = \sigma^2 < \infty$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$E[\bar{X}] = \mu, Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

5.5 The normal approximation to χ^2

$$Y \sim \chi^2(n), n = 1, 2, \cdots$$

$$Z_n = \frac{Y_n - n}{\sqrt{2n}} \to^d Z \sim N(0,1)$$

Recall: If $X_i \sim \chi^2(k_i)$ are independent, then $\sum_{i=1}^n X_i \sim \chi^2(\sum_{i=1}^n k_i)$.

From C.L.T.,

$$\frac{(\bar{X}_n - E[X_i])}{\sqrt{\frac{Var(X_i)}{n}}} \to^d Z \sim N(0,1)$$

where $E[X_i] = 1, Var(X_i) = 2$.

i.e.

$$\frac{\frac{Y_n}{n} - 1}{\sqrt{\frac{2}{n}}} \to^d Z \sim N(0,1)$$

i.e.

$$Z_n = \frac{Y_n - n}{\sqrt{2n}} \to^d Z \sim N(0,1)$$

5.6 Weak Law of Large Number (WLLN)

$X_1, X_2, \cdots, X_n, \cdots$ are independent and $E[X_i] = \mu, Var(X_i) = \sigma^2 < \infty$.

$$\bar{X}_n \to^P \mu$$

Proof. $\bar{X}_n \to^P \mu$, for any $\epsilon > 0$.

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

By Markov Inequality,

$$P(|\bar{X}_n - \mu| > \epsilon) \le \frac{E(|\bar{X}_n - \mu|^k)}{\epsilon^k}$$

Let $k = 2$, Then

$$P(|\bar{X}_n - \mu| > \epsilon) \le \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

hemce $\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$ □

5.6.1 Example

$X_i \sim Unif(0, \theta)$ are independent and $Y_n = \max\{X_1, \cdots, X_n\}$. The pdf for X_i is

$$f_i(x) = \frac{1}{\theta}, 0 < x < \theta$$

the cdf of X_i is

$$F_i(x) = \int_0^x \frac{1}{\theta} dt = \frac{x}{\theta}, x \in (0, \theta)$$

If $x \leq \theta$, $F_i(x) = 0$; if $x \geq \theta$, $F_i(x) = 1$.

$$H_n(y) = P[Y_n \leq y] = P[X_1 \leq y, X_2 \leq y, \cdots, X_n \leq y] = \prod_{i=1}^{n} P[X_i \leq y]$$

(If $y \leq 0, H_n(y) = 0$; if $y \geq \theta, H_n(y) = 1$)

If $y \in (0, \theta)$,

$$H_n(y) = \prod_{i=1}^{n} [\frac{y}{\theta}] = (\frac{y}{\theta})^n$$

Hence

$$\lim_{n\to\infty} H_n(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ 0 & \text{if } 0 < y < \theta \\ 1 & \text{if } y \geq \theta \end{cases}$$

i.e. $Y_n \to^P \theta (Y_n \to^d \theta)$

5.6.2 Example

$$X_i \sim^{iid} Exp(1, \theta)$$

$$f_i(x) = e^{-(x-\theta)}, x \geq \theta$$

$$Y_n = \max(X_1, X_2, \cdots, X_n)$$

$Y_n \to^d \theta$ and $Y_n \to^P \theta$.

$$V_n = n(Y_n - \theta)$$

$$G_n(V) = P(V_n \leq v) = P[Y_n \leq \frac{V}{n} + \theta]$$

(Recall: the cdf of Y_n is $1 - e^{-n(y-\theta)}$, for $y > \theta$)

$$G_n(V) = F_n(\frac{V}{n} + \theta) = 1 - e^{-n(\frac{v}{n} + \theta - \theta)} = 1 - e^{-v}, \text{for } v > 0$$

$$V_n \sim G_n(v) = 1 - e^{-v}, v > 0$$

Hence $V_n \sim Exp(1)$.

$W_n = n^2(Y_n - \theta)$, W_n does not have a limiting distribution. (The limit of cdf's of W_n is not a cdf)

5.6.3 Example

$X_1, \cdots, X_n, \cdots \sim^{iid} Gamma(2, \theta)$

$$E[X_i] = 2\theta, Var(X_i) = 2\theta^2$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

From WLLN,

$$\bar{X}_n \to^P E[X_i] = 2\theta$$

$(\bar{X}_n \to^d 2\theta)$ Apply CLT, then we done.

5.7 The limit Theorems

1. If $X_n \to^P a$ and $g(\cdot)$ is continuous at $x = a$, then $g(X_n) \to^P g(a)$. $X_1, \cdots, X_n, \cdots$, $X_n \to^P 2$, $g(x) = x^2$ and $g(\cdot)$ is continuous at $x = 2$. Hence, $g(X_n) \to^P g(2)$.

2. $X_n \to^P a$ and $Y_n \to^P b$. and $g(x, y)$ is continuous at (a, b). Hence $g(X_n, Y_n) \to^P g(a, b)$.

3. $\begin{cases} X_n \to^d X \\ Y_n \to^P b \end{cases}$, $g(x, b)$ is continuous at x in the support set of X

$$g(X_n, Y_n) \to^d g(x, b)$$

For example, $X_1, X_2, \cdots, X_n, \cdots, X_n \to^d Z \sim N(0, 1)$.

$Y_1, \cdots, Y_n, \cdots$, $Y_n \to^P 2$.

$g(x, y) = xy$ and $g(X_n, Y_n) \to^d g(z, 2)$. Since $g(x, 2) = 2x$ is continuous at $x \in \mathbb{R} = \mathbb{R}_x$.

Then $g(z, 2) = 2z$ as $z \sim N(0, 1)$ and $g(z, 2) = 2z \sim N(0, 4)$.

Here is another example:

$$\begin{cases} X_n \to^P a(> 0) \\ Y_n \to^P b(\neq 0) \\ Z_n \to^d Z \sim N(0, 1) \end{cases}$$

Firstly, For X_n^2, $X_n \to^P a$, $g(x) = x^2$ which is continuous everywhere (at a). Hence

$$g(X_n) \to^P g(a) = a^2$$

Secondly, for $\sqrt{X_n}$, $X_n \to^P a$, $g(x) = \sqrt{x}$, is continuous at $a > 0$,

$$g(X_n)^P \to^P g(a) = \sqrt{a}$$

Thirdly, $X_n Y_n$, $g(x, y) = xy$ is continuous at (a, b). Therefore,

$$g(X_n, Y_n) \to^P g(a, b) = ab$$

Next, $2Z_n$, with $\lim_{n\to\infty} a_n = 2$,

$$g(Z_n, a_n) = 2Z_n \to^d g(Z, 2)$$

Lastly, $X_n Z_n$, $g(x, y) = xy$,

$$g(X_n, Z_n) \to^d g(a, z) = az \sim N(0, a^2)$$

5.8 Delta Method

6 Maximum Likelihood Estimation

6.1 Introduction

$X_1, \cdots, X_n$ are i.i.d. with $f(x, \theta)$, $\theta \in \Omega$. For example, $X_i \sim^{iid} Pois(\theta), \Omega = \mathbb{R}^+$. Estimate θ.

$$X = (X_1, \cdots, X_n)$$

For example, to estimate θ is $Pois(\theta)$ use $\bar{X}$, which is a statistic (defined as a function of observable data), does not depend on unknown parameters. Another example will be the $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. On the other hand, however, $S_0^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \mu)^2$ where $\mu = E[X_i]$ which is unknown!. Hence it is not a statistic. Also, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is not a statistic.

Estimators $T(X) = T(X_1, \cdots, X_n)$ is used to estimate $\tau(\theta)$.

Estimate If $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$, then $T(x) = T(x_1, x_2, \cdots x_n)$.

For example, $X = (X_1, \cdots, X_n)$ where $X_i \sim^{iid} Pois(\theta)$. $T(x) = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ is an estimator of $\tau(\theta) = \theta$. If $n = 5, X_1 = 3, X_2 = 2, X_3 = 1, X_4 = 5, X_5 = 6$, estimate of θ is 3.4.

6.2 Maximum Likelihood Estimation - One Parameter

"Likelihood" is a probability $X \sim f(x;\theta)$ a probability function where $f(x,\theta) = P[X = x;\theta]$ is likelihood of θ.

Likelihood function is $L(\theta) = P[\text{observing the data};\theta]$. Suppose $X = (X_1, \cdots, X_n)$, $X_i \sim^{iid} f(x;\theta)$. Then $L(\theta) = \prod_{i=1}^{n} P[X_i;\theta] = \prod_{i=1}^{n} f(x;\theta)$ where $X_i = x_i$. Try to maximize $L(\theta)$ to find estimator of $\tau(\theta) = \theta$. Take the log-likelihood

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^{n} \log(f(x;\theta))$$

For example, $X_1, \cdots, X_n \sim^{iid} Ber(\theta)$. We observed x successes.

$$L(\theta) = P(\text{observing x successes};\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

Maximum Likelihood Estimate $\hat{\theta} = \hat{\theta}(x) =$ any max $l(\theta)$. or $\hat{\theta} = \hat{\theta}(x) =$ any max $L(\theta)$.

Maximum Likelihood Estimator $\hat{\theta} = \hat{\theta}(x) =$ any max $L(\theta;x)$ or $\hat{\theta} = \hat{\theta}(x) =$ any max $l(\theta;x)$

6.2.1 Example

"x" successes in a sequence of "n" Bernoulli trials with success probability θ. $L(\theta) = L(\theta;x) = P[\text{observing x successes in "n" Bernoulli trials};\theta] = \binom{n}{x}\theta^x(1-\theta)^{n-x}$. Hence

$$l(\theta) = l(\theta;x) = \log(L(\theta;x)) = x\log\theta + (n-x)\log(1-\theta) + \log\binom{n}{x}$$

$$\frac{d}{d\theta}l(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0$$

Therefore, $\hat{\theta} = \hat{\theta}(x) = \frac{x}{n}$. To test it is true, we need to check

$$\frac{d}{d\theta}l(\theta) > 0 \text{ for } \theta < \hat{\theta}$$

$$\frac{d}{d\theta}l(\theta) < 0 \text{ for } \theta > \hat{\theta}$$

Therefore, $\frac{x}{n}$ is the ML estimate; $\frac{X}{n}$ is the ML estimator.

6.2.2 Example

$X_1, \cdots, X_n \sim^{iid} Poi(\theta)$. We observe "$x_1, \cdots, x_n$".

$$L(\theta) = L(\theta; x) = P(\text{observing } x_1, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

.

$$\frac{d}{d\theta} l(\theta) = -n + \frac{\sum_{i=1}^{n} x_i}{\theta} + 0 = 0$$

Hence $\hat{\theta} = \hat{\theta}(x) = \bar{x}$ where $x = (x_1, \cdots, x_n)$.

Check the first order and the second order to verify, it is the ML estimate.

6.3 Information Function

Score Function $S(\theta) = l'(\theta)$

Information Function $-l''(\theta)$

If you are trying to find the sampling property of $l(\theta) = l(\theta; X)$.

Likelihood function for continuous models $X_1, X_2, \cdots, X_n \sim^{iid} f(x; \theta)$.

$$L(\theta) = \prod_{i=1}^{n} f(x; \theta)$$

$$S(\theta) = l'(\theta), S(\theta; x)$$

$$I(\theta) = -l''(\theta)$$

$$I(\hat{\theta}) = I(\theta)|_{\theta = \hat{\theta}(x)}$$

$$J(\theta) = E[I(\theta; x)]$$

6.3.1 Example

$X_1, \cdots, X_n \sim^{iid} Unif(0, \theta)$, Find ML estimator of θ.

$$L(\theta) = \prod_{i=1}^{n} f(x; \theta) = \frac{1}{\theta^n} I[0 \leq \max\{x_1, \cdots, x_n\} \leq \theta]$$

Therefore, $\hat{\theta} = \max\{x_1, \cdots, x_n\}$.

6.3.2 Example

$X_1, \cdots, X_n \sim^{iid} Wei(1, \theta)$. Given

$$f(x;\theta) = \theta x^{\theta-1} e^{-x^\theta}, x > 0, \theta > 0$$

$$L(\theta) = \prod_{i=1}^{n} f(x;\theta) = \theta^n [\prod_{i=1}^{n} x_i]^{\theta-1} e^{-\sum_{i=1}^{n} x_i^\theta}$$

$$l(\theta) = n \log\theta + (\theta - 1)\sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i^\theta$$

$$S(\theta) = l'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i = -\sum_{i=1}^{n} \log x_i x_i^\theta$$

We cannot easily find the value of this score function so we need some numerical method to work out the approximate solution.

6.3.3 Newton's Method

For solving $f(x) = 0$, $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$

We can use the same method so that $\theta^{(i+1)} = \theta^{(i)} - \frac{S(\theta^{(i)}}{S'(\theta^{(i)})}$. Hence $\theta^{(i+1)} = \theta^{(i)} + \frac{S(\theta^{(i)}}{I(\theta^{(i)})}$ and stop if $|\theta^{(i+1)} - \theta^{(i)}| < e^{-10}, S(\theta^{(i+1)}) \approx 0$.Well we need to verify that $\theta^{(i+1)}$ is maximizing $L(\theta)$.

$\hat{\theta}$ is the ML estimate of θ. To find the ML estimate of $\tau(\theta)$, a one-to-one function! $\Longrightarrow$ $\hat{\tau}(\theta) = \tau(\hat{\theta})$ by the invariance property in MLE. For example, $X_1, \cdots, X_n \sim^{iid} Poi(\mu)$, the ML estimator of μ

$$\hat{\mu} = \bar{x}$$

Let $\tau(\mu) = \mu^2$ which is one-to-one for $\mu > 0$. Then the ML estimate of $\tau(\mu)$

$$\hat{\tau}(\theta) = \tau(\hat{\theta}) = (\bar{x})^2$$

$$f(x;\theta) = \theta x^{\theta-1} e^{-x^\theta}, x > 0, \theta$$

$$L(\theta) = \theta^n (\prod_{i=1}^{n} x_i)^{\theta-1} e^{-\sum_{i=1}^{n} x_i^\theta}$$

$$l(\theta) = n \log\theta + (\theta - 1)\sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i^\theta$$

$$S(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i - \sum_{i=1} (\log x_i) x_i^\theta$$

$$I(\theta) = (-1)[\frac{-n}{\theta^2} - \sum_{i=1}^{n}(\log x_i)^2 x_i^\theta]$$

That is

$$I(\theta) = \frac{n}{\theta^2} + \sum_{i=1}^{n}(\log x_i)^2 x_i^\theta$$

To use the Newton's Method to find the M.L. estimate.

Step 1 Initial $\theta^{(0)}$ estimated from data (graphing $L(\theta)/l(\theta)$).

Step 2

$$\theta^{(i+1)} = \theta^{(i)} + \frac{\frac{n}{\theta^{(i)}} + \sum_{j=1}^{n}\log x_j - \sum_{i=1}^{n}(\log x_i)x_i^{\theta^{(i)}}}{\frac{n}{[\theta^{(i)}]^2} + \sum_{j=1}^{n}(\log x_j)^2 x_i^{\theta^{(i)}}}$$

until $|\theta^{(i+1)} - \theta^{(i)}| < e^{-10}$

Step 3 : $\hat{\theta} = \theta^{(i+1)}$, check if it is the ML estimate.

Invariance of the ML estimator: if $\hat{\theta}$ is the ML estimator of θ. $\tau(\hat{\theta})$ is the ML estimator of $\tau(\theta)$, which is a one-to-one function of θ.

6.3.4 Example

The ML estimate of θ is $\hat{\theta} = \frac{n}{\sum_{i=1}^{n}\log x_i}$. Find $\tau(\theta) =$ media. That means $P[X \leq \tau] = F(\tau) = 1/2$.

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \int_0^x \theta t^{\theta-1} dt & 0 < x < 1 \\ 1 & x \geq 1 \end{cases} = \begin{cases} 0 & x \leq 0 \\ x^\theta & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

$F(\tau) = 1/2, \tau \in (0,1)$ and then $F(\tau) = \tau^\theta = 1/2$. That is $\tau = \frac{1}{2}^{\frac{1}{\theta}}, \theta > 0$. $\tau(\theta) = \frac{1}{2}^{\frac{1}{\theta}}$ is a one-to-one function of θ. The ML estimator of $\tau(\theta)$ is $\tau(\hat{\theta}(x)) = \frac{1}{2}^{\frac{1}{\hat{\theta}(x)}}$.

6.4 Likelihood Region

$f(x;\theta), \theta \in \Omega$ and $\hat{\theta}$ are kind of point-estimate. Define

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}$$

$L(\hat{\theta}) = L(\theta)|_{\theta=\hat{\theta}}$ where $\hat{\theta}$ is the ML estimate.

$$\hat{\theta} = arg\max_{\theta} L(\theta)$$

That is $0 \leq R(\theta) \leq 1$.

For example, $P = 0.15 = 100 \cdot p\%$. The likelihood region pf θ is $\{\theta : R(\theta) \geq p = 0.15\}$.

50% likelihood region, that is, $p = 0.5$, $\theta : R(\theta) \geq 0.5\}$.

Example: $X_1, \cdots, X_n \sim^{iid} Poi(\theta)$. $R = 30.\hat{\theta} = 5$. 10% likelihood Region is $\{\theta : R(\theta) \geq 0\}$. $R(\theta) = \frac{L(\theta))}{L(\hat{\theta})} = \frac{\prod f(x;\theta)}{\prod f(x;\hat{\theta})}, \hat{\theta} = \bar{\theta}$. Therefore, it becomes

$$R(\theta) = \frac{e^{-n\theta+n\hat{\theta}}\theta^{\sum x_i}}{\hat{\theta}^{\sum x_i}} = \frac{e^{-30(\theta-5)}\theta^{150}}{5^{150}}$$

10% likelihood region is $\{\theta : R(\theta) \geq 0.1\}$.

Let $R(\theta) = P$. (Recall Newton's Method is for solving $f(x) = 0$). That is $f(\theta) = R(\theta) - P = 0$.

$$[I\hat{\theta}]^{1/2}(\hat{\theta}_n - \theta_0) \to^d Z \sim N(0,1) \implies \hat{\theta}_n \sim N(\theta_0, \frac{1}{I(\hat{\theta})})$$

$$[J\hat{\theta}]^{1/2}(\hat{\theta}_n - \theta_0) \to^d Z \sim N(0,1) \implies \hat{\theta}_n \sim N(\theta_0, \frac{1}{J(\hat{\theta})})$$

6.5

$X = \{X_1, \cdots, X_n\}, X_i \sim^{iid} f(x;\theta), \theta \in \Omega$. $(A(x), B(x))$ is an interval estimate. Confidence interval statistic is an interval estimator. 100p% LR for θ $\{\theta : R(\theta \geq P\}$ For example, 10% LR $= \{\theta : R(\theta) \geq 0.1\}$. Pivotal Quantity: $Q(X;\theta)$ is not a statistic!. Its distribution does not dependent on θ. $X = (X_1, \cdots, X_n), X_i \sim^{iid} N(0,1)$. To find a Pivotal quantity!. $\bar{X} \sim N(\mu, \frac{1}{n})$. Therefore $Q(x;\mu) = \sqrt{n}(\bar{X} - \mu) \sim N(0,1)$.

Recall: $\bar{X} \sim N(\mu, \frac{1}{n}$ and $(\bar{X} - \mu) \sim N(0, \frac{1}{n})$, $\sqrt{n}(\bar{X} - \mu) \sim \sqrt{n}N(0, \frac{1}{n}) = N(0,1)$. $Q(X,\mu) = \sqrt{n}(\bar{X} - \mu) \sim N(0,1)$. That is, $Q(X,\mu)$ is a Pivotal quantity!.

For example, $X = (X_1, X_2, \cdots, X_n), X_i \sim^{iid} N(0,\sigma^2), Q(X;\sigma^2)$. Therefore $\bar{X} \sim N(0, \sigma^2/n)$. $Q(X,\sigma^2) = \frac{\sqrt{n}\bar{X}}{\sqrt{\sigma^2}} \sim N(0,1)$. Therefore $\frac{n\bar{X}^2}{\sigma^2} \sim \chi^2(1)$.

If the limiting distribution of $Q(X,\theta)$ does not dependent on θ, it is an asymptotic Pivotal quantity.

For example, $X = (X_1, \cdots, X_n), X_i \sim^{iid} Poi(\theta)$. $\frac{\sqrt{n}(\bar{X}_n - \theta)}{\bar{X}_n}$

Due to WLLN, we know $\bar{X}_n \to^P E[X_i] = \theta$. $\frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\theta}} \to^d Z \sim N(0,1)$. Due to CLT, $E[X_i] = \theta, Var(X_i) = \theta$, $\frac{\sqrt{n}(\bar{X}_n-\theta)}{\sqrt{\theta}} \to^d Z \sim N(0,1)$ and $\bar{X}_n \to^P \theta$. $\sqrt{n}[\frac{\bar{X}_n-\theta}{\sqrt{\bar{X}_n}}] = [\sqrt{n}\frac{(\bar{X}_n-\theta)}{\sqrt{\theta}}][\frac{\sqrt{\theta}}{\sqrt{\bar{X}_n}}]$.

Due to Slutscky Theorem, we know that it is convergent to distribution $N(0,1)$. $Q(X,\theta)$ is an asymptotic Pivotal Quantity!.

6.5.1 Confidence Interval

If $P(A(x) \leq \theta \leq B(x)) = P, 0 < p < 1$. $(A(x), B(x))$ is a 100 p% CI for θ. If $X = (X_1, \cdots, X_n)$ is a random sample. If $p = 0.95$, out of 1000 confidence intervals about 950 will cover θ_0. 100p% CI for θ is $P[q_1 \leq Q(X;\theta) \leq q_2] = p$. Therefore $P(A(x) \leq \theta \leq B(x)] = p$.

Example 1 $\sqrt{n}(\bar{X} - \mu) \sim N(0,1)$, 95% CI for μ. $Q(\bar{X}, \mu) = \sqrt{n}(\bar{X} - \mu)$ is a monotone function of μ. $P[q_1 \leq Q(x,\mu) \leq q_2) = 0.95$. (equal tail CI $P(Q(X;\mu) < q_1] = P[Q(x,\mu) > q_2]$. $q_1 = Z_{0.025}, q_2 = Z_{0.975}$.

$P[-1.96 \leq \sqrt{n}(\bar{X}_n - \mu) \leq 1.96] = 0.95$. Therefore, $P(\frac{1}{\sqrt{n}}(-1.96) - \bar{X}_n \leq -\mu \leq \frac{1}{\sqrt{n}}(1.96) - \bar{X}_n)$.

Example 2: $X_i \sim^{iid} Exp(\theta)$, $X_1 = (X_1, \cdots, X_n)$, then $\hat{\theta}(X) = \bar{X}_n$ is the ML estimator of θ. $Q = \frac{2n\hat{\theta}_n}{\theta} = \sum_{i=1}^{n}[\frac{2X_i}{\theta}]$. Since $X_n \sim^{iid} Exp(\theta)$ then $M_{X_i}(t) = \frac{1}{1-\theta t}, t \leq \frac{1}{\theta}$. $Y_i = \frac{2X_i}{\theta}$, mgf of Y_i is $M_{Y_i}(t) = M_{X_i}(\frac{2t}{\theta})$.

$$M_Q(t) = \prod_{i=1}^{n} M_{Y_i}(t) = \prod_{i=1}^{n}[\frac{1}{1-\theta\frac{2t}{\theta}}] = \frac{1}{(1-2t)^2} t < \frac{1/\theta}{2/\theta} = \frac{1}{2}$$

$\chi^2(2n)$'s mgf is $\frac{1}{(1-2t)^n}, t < \frac{1}{2}$. Therefore, $Q \sim \chi^2(2n)$ is a Pivotal quantity.

$$Q = \frac{2\sum X_i}{\theta}$$

monotonically decreasing in θ

$$\begin{aligned} P(q_1 \leq Q \leq q_2) = 0.95 &= P(\chi^2_{0.025}(2n) \leq Q \leq \chi^2_{0.9725}(2n)) = 0.95 \\ &= P[(2n\bar{X}_n)(\frac{1}{\chi^2_{0.975}(2n)}) \leq \theta \leq (2n\bar{X}_n)(\frac{1}{\chi^2_{0.025}(2n)})) = 0.95 \end{aligned}$$

When $n = 15$, $\sum_{i=1}^{n} x_i = 36$, 95% equal tail C.I. for θ is $((2n\bar{X}_n)(\frac{1}{\chi^2_{0.975}(2n)}), (2n\bar{X}_n)(\frac{1}{\chi^2_{0.025}(2n)}))$

Example 3: $X = (X_1, \cdots, X_n), X_i \sim^{iid} Exp(1,\theta)$. $f(x,\theta) = e^{-(x-\theta)}, x \geq \theta, \theta > 0$, $L(\theta) = \prod_{i=1}^{n} e^{-(x_i-\theta)} = [e^{-\sum_{i=1}^{n} x_i}][e^{n\theta}]$. To maximize $L(\theta)$, we are to find the largest θ. If we observe $X_1 = x_1, \cdots, X_n = x_n$, and $X_{(n)} \geq \cdots \geq x_{(1)} \geq \theta$. ML estimate: $\hat{\theta} = X_{(1)}$, the smallest observation. ML estimate is $\hat{\theta}(x) = X_{(1)}$. $P[\hat{\theta} - \theta \leq q] = 1 - e^{-nq}, q \geq 0$ is the cdc of $\hat{\theta} - \theta$. Then $(\hat{\theta} - \theta)$ is a pivotal quantity!

$P[\hat{\theta} - \theta \leq q] = P[X_{(1)} - \theta \leq q] = P[X_{(1)} \leq \theta + q] = 1 - P[X_{(1)} > \theta + q] = 1 - \prod P[X_i > \theta + q] = 1 - \prod[1 - F_i(\theta + q)]$ where $F_i(\cdot)$ is the cdf of $X_i \sim Exp(1, \theta + q)$. Hence, it is $1 - \prod_{i=1}^{n}[1 - (1 - e^{-(q+\theta-\theta)}) = 1 - e^{-nq}$.

Is $[\hat{\theta} + n^{-1}\log(1-p), \hat{\theta}]$ 100p%? To verify this we need to find

$$\begin{aligned} P[\hat{\theta} + n^{-1}\log(1-p) \le \theta \le \hat{\theta}] = p = P[n^{-1}\log(1-p) \le \theta - \hat{\theta} \le 0] \\ = P[(-1)n^{-1}\log(1-p) \ge (-1)(\theta - \hat{\theta}) \ge (-1)\theta] \\ = P[0 \le (\hat{\theta} - \theta) \le \frac{-\log(1-p)}{n}] = G(-\frac{\log(1-p)}{n}) - G(0) = p \end{aligned}$$

If

$$\begin{aligned} P[\hat{\theta} + n^{-1}\log(\frac{1-p}{2}) \le \theta \le \hat{\theta} + n^{-1}\log(\frac{1+p}{2})] \overset{?}{=} p \\ = P[n^{-1}\log(\frac{1-p}{2}) \le \theta - \hat{\theta} \le n^{-1}\log(\frac{1+p}{2})] \\ = G(n^{-1}\log(\frac{1+p}{2})) - G(-n^{-1}\log(\frac{1+p}{2})) \\ = (1 - (\frac{1-p}{2})) - (1 - (\frac{1-p}{2})) = p \end{aligned}$$

The second one, $\hat{\theta}$ is not in this C.I.. The first C.I. is more reasonable.

$X = (X_1, \cdots, X_n)$, $Q(x;\theta)$ is asymptotic Pivotal quantity if the limiting distribution of $\theta(x,\theta)$ does not depend on θ. $[J(\hat{\theta}_n)]^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \to^d Z \sim N(0,1)$. $P[q_1 \le [J(\hat{\theta}_n)]^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \le q_2] \approx p = P[A(x) \le \theta_0 \le B(x)] \approx p$. If approximate 90% equal tail C.I. for θ_0, $q_1 = Z_{0.05}$ and $q_2 = Z_{0.95}$. $P[Z_{0.05} \le [J(\hat{\theta}_n)]^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \le Z_{0.95}] = 0.9 = P[\frac{Z_{0.05}}{[J(\hat{\theta}_n)]^{\frac{1}{2}}} \le \theta_0 \le \hat{\theta}_n - \frac{Z_{0.95}}{[J(\hat{\theta}_n)]^{\frac{1}{2}}}] \approx 0.9$

$$P[-a < Z < a] = 0.9$$

where $a = Z_{0.95}$ and$-a = Z_{0.05}$.

In summary, an approximate 100p% C.I. with equal tails is $[\hat{\theta}_n - a[J(\hat{\theta}_n)]^{-\frac{1}{2}}, \hat{\theta}_n + a[J(\hat{\theta})]^{-\frac{1}{2}}]$.

Example: $X \sim Bin(n,\theta)$. Find an approximate 95% equal tail C.I. for θ is

$$\hat{\theta}_n = \frac{X}{n}$$

$$L(\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$l(\theta) = \log(\binom{n}{x} + x\log(\theta) + (n-x)\log(1-\theta)$$

$$S(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

$$I(\theta) = -S'(\theta) = -l''(\theta) = \frac{n\theta(1-\theta) + (x-n\theta)(1-2\theta)}{\theta^2(1-\theta)}$$

$$J(\theta) = E[I(\theta; x)] = \frac{n}{\theta(1-\theta)}$$

as $E[X] = n\theta$. $[\hat{\theta}_n - a[J(\hat{\theta})]^{-\frac{1}{2}}, \hat{\theta}_n + a[J(\hat{\theta}_n)]^{-\frac{1}{2}}] = [\frac{x}{n} - 1.96[\frac{n}{\frac{x}{n}(1-\frac{x}{n})}]^{-\frac{1}{2}}, \frac{x}{n} + 1.96[\frac{n}{\frac{x}{n}(1-\frac{x}{n})}]^{-\frac{1}{2}}]$

$$(I(\hat{\theta}_n))^{-\frac{1}{2}}(\hat{\theta} - \theta_0) \to^d Z \sim N(0,1)$$

An approximate 100p% equal tail C.I. for θ_0

$$[\hat{\theta}_n - a[I(\hat{\theta}_n)]^{-\frac{1}{2}}, \hat{\theta}_n + a[I(\hat{\theta}_n)]^{-\frac{1}{2}}$$