## Distribution

$X$ and $Y$ are rvs with joint pdf $p_{XY}(x,y)$, then $p_{Y|X}(y|x) = P(Y=y|X=x) = \frac{P(X=x,Y=y)}{P(X=x)} = \frac{p_{XY}(x,y)}{p_X(x)}$ Function of rvs:
If $X, Y$ are differentiable, 1 to 1 transfromation, and $f_X(x)$ is the pdf of $X$, and $Y=g(X)$, where $g$ is differentiable and strictly monotone, then

$$f_Y(y) = f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}|$$

If $X=(X_1,..,X_n)$ is multivariate continuous and $Y=(Y_1,\dots,Y_n) = g(X)$ where $g$ is 1 to 1 i.e. $X=g^{-1}(Y)=w(Y)$ assume $w$ have continuous partial derivatives, and let

$$J = \begin{vmatrix} \frac{\partial w_1(y)}{\partial y_1} & \dots & \frac{\partial w_1(y)}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial w_n(y)}{\partial y_1} & \dots & \frac{\partial w_n(y)}{\partial y_n} \end{vmatrix}$$

then $f_Y(y)=f_X(g^{-1}(y))|J|$
**Example**: $Y=X_2/X_1$. Let $Y_1=\frac{X_2}{X_1}, Y_2=X_1$, then $X_2=Y_1Y_2$, then $w_1(y)=y_2, w_2(y)=y_1y_2$, then $J=\begin{bmatrix}0 & 1\\ y_2 & y_1\end{bmatrix}$ and

$$f_Y(y) = f_X(x)|J| = f_X(y_2,y_1y_2)|J| = f_X(y_2,y_1y_2)|-y_2|$$

the marginal pdf of $y_1$ is

$$f_{Y_1}(y)=\int_{-\infty}^{\infty} f_X(y_2,y_1y_2)|-y_2|\,dy_2 = \int_{-\infty}^{\infty}|x_1|f_X(x_1,x_1y)\,dx_1$$

since $y_2=x_1$
for $Z=X+Y$
$f_Z(z)=\int_{-\infty}^{\infty} f_X(z-y)f_Y(y)\,dy = \int_{-\infty}^{\infty} f_Y(z-x)f_X(x)\,dx$

## Moment

$M_X(t)=\Phi(s)=E(e^{tX})=\int_{-\infty}^{\infty} f(x)e^{tx}\,dx)$
Properties:
1. $M_X^{(r)}(t)=\mathbb{E}\left[\frac{d^r}{dt^r}e^{tX}\right]=\mathbb{E}[X^re^{tX}]$ $\quad M_X^{(r)}(0)=\mathbb{E}[X^r]$
2. if $Y=aX+b$, then $M_Y(t)=e^{bt}M_X(at)$

## E, Var, Cov

Under condition $X$, the mean of $Y$ is $E(Y|X)$. Then we compute mean respect to $X$, we get the mean of $Y$. Which means

$$E(Y)=E_X(E_Y(Y|X))$$

Then

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

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

add the two equations

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

$$\mathrm{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))] \quad = E(XY)-E(X)E(Y)$$

## Sample mean and Var

let $X_1,\dots,X_n$ be iid such that $X_i\sim N(\mu,\sigma^2)$, define
- Sample Mean: $\bar X = \frac1n\sum X_i$
- Sample Variance: $S_n^2=\frac{1}{n-1}\sum(X_i-\bar X)^2$

Then we have
1. $\bar X\sim N\left(\mu,\frac{\sigma^2}{n}\right)$ and $\sqrt n(\bar X-\mu)/\sigma\sim N(0,1)$
2. $\bar X, S_n^2$ are independent
3. $(n-1)S_n^2/\sigma^2\sim\chi^2_{n-1}$
4. $\sqrt n(\bar X-\mu)/S_n\sim t_{n-1}$

To calculate $I(f)=\int_0^1 f(x)\,dx$, we can generate $X_1,\dots,X_n$ iid and $\sim U(0,1)$ and compute

$$I(f)=\frac1n\sum f(X_i)$$

By LLN, $\hat I(f)$ will be close to $\mathbb{E}[f(X)]=\int_0^1 f(x)\cdot 1\,dx = I(f)$ as $n$ is large enough
Let $\bar X_n$ be the sample average and $T_n=n\bar X_n$ be the sum of data, then

$$\lim_{n\to\infty} P\left(\frac{T_n-n\mu}{\sigma\sqrt n}\le x\right)=\lim_{n\to\infty} P\left(\frac{\sqrt n(\bar X_n-\mu)}{\sigma}\le x\right)=\Phi(x)$$

## Common Distri

**Uniform Distri**:
pdf: $f(x)=\frac{1}{b-a}$, cdf: $F(x)=\frac{x-a}{b-a}$ $\quad a\le x\le b$
mgf: $M_X(t)=\mathbb{E}[e^{tX}]=\int\frac{1}{b-a}e^{tX}\,dx=\frac{e^{tb}-e^{ta}}{t(b-a)}$
If $Y$ is a rv with cdf $F_Y(y)$, we let $W=F_Y(y)$, then the cdf of $W$ is

$$F_W(w)=P(W\le w)=P(F_Y(y)\le w)$$

since $F_Y(y)$ is strictly increasing, the it has inverse function, thus

$$P(W\le w)=P(y\le F_Y^{-1}(w))=F_Y(F_Y^{-1}(w))=w$$

which means $W\sim U(0,1)$ this imply that any continouse rv can be transformed into uniform distri
**Bernoulli Distribution**, $X\sim B(p)$:
pmf: $p(x)=p^x(1-p)^{1-x}$ $\quad x=0$ or $1$
mgf: $M_X(t)=\mathbb{E}[e^{tX}]=pe^{1t}+(1-p)e^{0t}=pe^t+1-p$
**Binomial Distribution**, $X\sim B(n,p)$:
pmf: $P(X=k)=C_n^kp^kq^{n-k}$
mgf: $M_X(t)=\prod_{i=1}^n(pe^t+1-p)=(pe^t+1-p)^n$
$\mathbb{E}[X]=np$, $\mathrm{Var}[X]=np(1-p)$
**Poisson Distri**, $X\sim P(\lambda)$:
pmf: $P(X=k)=\pi_k(\lambda)=\frac{e^{-\lambda}}{k!}\lambda^k$, $k\in \mathrm{N}^0$, mean: $\lambda$, Var: $\lambda$
mgf: $M_X(t)=\mathbb{E}[e^{tX}]=e^{-\lambda}\sum\frac{\lambda^xe^{tx}}{x!}=e^{-\lambda}e^{\lambda e^t}=e^{\lambda(e^t-1)}$
If $X_i\sim P(\lambda_i)$, let $Y=X_1+\dots+X_k$, then $Y\sim P(\lambda_1+\lambda_2+\dots+\lambda_k)$
**Negative Binomial Distribution**, $X\sim NB(r,p)$:
$X$ is said to be negative binomial random variable with parameters $r$ and $p$ if

$$P(X=x)=C_{x-1}^{x-r}p^rq^{x-r}$$

where $k=r,r+1,\dots\infty$
The negative binomial distribution shows if we want to make the events occur $r$ times, what is the prob if we make exactly $k$ times trials
mgf: $M_X(t)=\frac{p^re^{rt}}{[1-(1-p)e^t]^r}$, mean: $r/p$, Var: $r(1-p)/p^2$
**Exponential Distribution**, $X\sim\exp(\lambda)$:
pdf: $f(x)=\lambda e^{-\lambda x}$, $x\ge0$ cdf: $F(x)=1-e^{-\lambda x}$, $x\ge0$
mean: $1/\lambda$, Var: $1/\lambda^2$, mgf: $M_X(t)=\frac{\lambda}{\lambda-t}$
The memoryless property:

$$P(X-y>x|X>y)=P(X>x),\ \text{for all } x\ge0 \text{ and } y\ge0$$

**Gamma Distribution**, $Y\sim\Gamma(\alpha,\lambda)$:
pdf: $f(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$, mgf: $M_X(t)=\left(\frac{\lambda}{\lambda-t}\right)^\alpha$
$\mathrm{E}=\alpha/\lambda$ $\quad \mathrm{Var}=\alpha/\lambda^2$ $\quad \Gamma(n)=(n-1)!$ $\quad \Gamma(1/2)=\sqrt\pi$
Normal Distribution:
pdf: $f(x)=\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac12\frac{(x-\mu)^2}{\sigma^2}}$, mgf: $e^{\mu t+\frac{\sigma^2t^2}{2}}$

## Sample Distri

**Chi-square Distribution**, $X\sim\chi_n^2$:
If rvs $X_1,\dots,X_n$ are iid and $X_i\sim N(0,1)$. Then their square sum is a new rv $X$ which follow Chi-square Distribution, denoted by $X\sim\chi^2_n$ or $X\sim\chi^2(n)$
pdf: $f(x)=\frac{1}{\Gamma(n/2)2^{n/2}}x^{\frac n2-1}e^{-\frac x2}$, mgf: $M_X(t)=\left(\frac{1}{1-2t}\right)^{\frac n2}$
mean: $n$, Var: $2n$, $\chi_n^2=\Gamma\left(\frac n2,\frac12\right)$
If $X_1,X_2,\dots,X_n$ are iid and $X_i\sim\chi^2_{n_i}$, let $Y=X_1+\dots+X_n$, then $Y\sim\chi^2_{n_1+\dots+n_n}$

$$M_Y(t)=\prod M_{X_i}(t)=\prod\left(\frac{1}{1-2t}\right)^{\frac{n_i}{2}}=\left(\frac{1}{1-2t}\right)^{\sum n_i/2}$$

**T Distribution**
If $X\sim N(0,1)$ and $Y\sim\chi^2_n$, then we say $T=X/\sqrt{Y/n}$ follows T Distribution, denoted by $T\sim t_n$
$\mathbb{E}[T]=0$, $\mathrm{Var}[T]=\frac{n}{n-2}$
**F Distribution**
Let $U\sim\chi^2_m, V\sim\chi^2_n$, be indepedent Chi-square Distribution, then F distribution is defined as

$$Y=\frac{U/m}{V/n}\sim F_{m,n}$$

Let $X\sim t_n$, then $Y=X^2\sim F_{1,n}$
$\mathbb{E}[Y]=\frac{n}{n-2}, n>2$, and $\mathrm{Var}[Y]=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$

## Estimation

A **point estimate** is a function $\hat\theta=g(X)$ of the observation vector/population $X=[x_1,\dots x_n]$. The corresponding rv $\hat\theta=g(X)$ is the point estimate of $\theta$
**Bias** of an estimator $\hat\theta$ is defined by

$$bias(\hat\theta)=\mathbb{E}[\hat\theta]-\theta$$

and we say $\hat\theta$ is an **unbiased estimator** of $\theta$ if $E(\hat\theta)=\theta$
**MSE** is defined by

$$MSE(\hat\theta)=\mathbb{E}[(\hat\theta-\theta)^2]=[bias(\hat\theta)]^2+\mathrm{Var}[\hat\theta]$$

If $\hat\theta-\theta$ tend to 0 in prob as $n\to\infty$, then $\hat\theta$ is called **consistent estimator**
The $k$th **moment** of the distribution is $\mu_k=\mathbb{E}[X^k]$
and the $k$th **sample moment** is defined by $m_k=\frac1n\sum X_i^k$
**Method of moments estimator (MME)** of the parameter(s) is obtained by solving the equation(s)

$$\hat\mu_k=m_k \qquad \text{for } k=1,\dots$$

the number of equations is equal to the unknown parameters
**Likelihood Function** $L(\theta|x)$ of parameter $\theta$ of random sample $X_1,\dots,X_n$ is defined by

$$L(\theta|\mathbf{x})=P(X_1=x_1,\dots,X_n=x_n)=\prod P(X_i=x_i|\theta)$$

for continuous:

$$L(\theta|\mathbf{x})=\prod f(x_i|\theta)$$

1.**Invariance**: Suppose that $\theta$ is the para associated with a distri. $g(\theta)$ is a function of $\theta$. If $\hat\theta$ is the MLE of $\theta$, then the MLE of $g(\theta)$ is given by $g(\hat\theta)$, i.e.

$$\hat\theta=\operatorname{argmax} L(\theta;X)$$

if $\phi=g(\theta)$, then the MLE of $\phi$ is

$$\hat\phi=g(\hat\theta)$$

2. **Consistency**: Under approiate smoothness conditions of pdf $f$, the MLE $\hat\theta$ is consistent (by Law of Large Numbers), i.e.

$$\hat\theta\xrightarrow{p}\theta \text{ as } n\to\infty$$

3. **Asymptotic Normality**: Under apprioate smoothness conditions of pdf $f$, the prob distri of $\sqrt{nI(\theta_0)}(\hat\theta-\theta_0)$ tends to standard normal (by Central Limit Theorem). That is

$$\sqrt{nI(\theta_0)}(\hat\theta-\theta_0)\sim N(0,1)$$

approximately for large $n$, where $\theta_0$ is the true value of para $\theta$. $I(\theta)$ is the Fisher Information
4. **Best Unbiased Estimator**: If the MLE $\hat\theta$ is unbiased, it would be the **best (minimum variance) unbiased estimator** (or **most efficient estimator**) when $n$ is large
Since the asymptotic variance of a MLE is equal to the lower bound, maximum likelihood estimates are said to be asymptotically efficient
**Fisher Information** is defined by

$$I(\theta)=\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]=-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right]$$

where $f(X|\theta)$ is the pdf of $X$ under given parameter $\theta$
**Cramer-Rao Lower Bound**: Let $X_1,\dots,X_n$ be iid sample with pdf $f(x|\theta)$ and $T=T(X_1,\dots,X_n)$ be an unbiased estimator of $\theta$. Then under appropriate smoothness conditions

$$\mathrm{Var}[T]\ge\frac{1}{nI(\theta)}$$

where $I(\theta)$ is Fisher Information

## confidence interval

Let $\mathbf{X}=(X_1,\dots,X_n)$ be a random sample with pdf $f(x|\theta)$, an interval estimator of $\theta$ with coverage prob $1-\alpha$ is a random interval $(\hat\theta_L(\mathbf{X}),\hat\theta_U(\mathbf{X}))$, where

$$P(\hat\theta_L(\mathbf{X})<\theta<\hat\theta_U(\mathbf{X}))=1-\alpha$$

the random interval is called an $100(1-\alpha)\%$ confidence interval $100(1-\alpha)\%$ is referred as **confidence level**
Let $Q(X,\theta)$ be a function of random sample. $Q(X,\theta)$ is a **pivot** if distri of $Q(X,\theta)$ is free of $\theta$ suppose $X_i\sim N(\mu,\sigma^2)$
1. if $\sigma^2$ is known, then $Q(X,\mu)=\frac{\sqrt n(\bar X-\mu)}{\sigma}\sim N(0,1)$ is pivot
2. if $\sigma^2$ is unknown, then $Q(X,\mu,\sigma)=\frac{\sqrt n(\bar X-\mu)}{S}\sim t_{n-1}$ is pivot
3. if $\sigma$ is known, then $Q(X,\mu)=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_n$ is pivot

Recall the Asymptotic Normality property. Then an approximate $100(1-\alpha)\%$ confidence interval is

$$1-\alpha\approx P\left(-z(\alpha/2)<\sqrt{nI(\hat\theta)}(\hat\theta-\theta_0)<z(\alpha/2)\right)$$

where $z(\alpha/2)$ is the upper $\alpha/2$ quantile in $N(0,1)$

## Sufficient Statistic

A Statistic $T(X_1,\dots,X_n)$ is said to be **sufficient** for parameter $\theta$ if the conditional distri of $X_1,\dots,X_n$ given $T=t$ dose not depend on $\theta$ for any value of $t$

**Factorization Theorem**: A necessary sufficient condition for $T(X_1,\dots,X_n)$ to be sufficient for a parameter $\theta$ is that the joint pdf or pmf factors in the form

$$f(x_1,\dots,x_n|\theta)=g(T(x_1,\dots,x_n),\theta)h(x_1,\dots,x_n)$$

If $T$ is sufficient for $\theta$, then the MLE for $\theta$ is a function of $T$

## Hypothesis Testing

We want to find a test to minimize the probability of Type II error, equivalently to maximize the **power**

$$power=1-P(\text{Type II Error})=1-\beta$$

The probability distribution of the test statistic when the **null hypothesis** is true, is referred to as **Null Distribution**

If null and alternative hypotheses each completely specify the probability distribution, then we call them **simple hypotheses**, otherwise it's **composite** (parameter is not a certain value)

**Neyman-Pearson Lemma/Paradigm**:

Considering Hypothesis Testing $H_0:\theta=\theta_0, H_1:\theta=\theta_1$ where the pdf or pmf corresponding to $\theta_i$ is $f(x|\theta_i), i=0,1$. Then the **Most Powerful** Test is to reject $H_0$ if $\Lambda(x)<C$ where $\Lambda(x)=f(x|\theta_0)/f(x|\theta_1)$ is the Likelihood Ratio and constant $C>0$. Furthermore the size of the test is given by

$$\alpha=\int_R f(x|\theta_0)\,dx$$

where $R=\{x:\Lambda(x)<C\}$

That is to say, if the likelihood ratio is small to some threshold $C$, then we reject $H_0$. When $\Lambda(x)<C$, we can solve it for $x$ and know the region where type I error happen, then we integral in this area to compute the the prob $\alpha$ of type I error happening. In practice, we first choose $\alpha$ and then determine $C$

In **Neyman-Pearson Paradigm**, we ask $H_0,H_1$ to be **simple hypotheses**. In practice, we often consider the testing $H_0:\theta=\theta_0, H_1=\theta>\theta_0$, where $H_1$ is a composite hypothesis

If a test of $H_0:\theta=\theta_0$ vs $H_1:\theta=\theta_1$ is a most powerful test for every $\theta_1>\theta_0$ among all test size of $\alpha$, then the test is **uniformly most powerful**, UMP for testing $H_0:\theta=\theta_0$ vs $H_1:\theta>\theta_0$

The alternatives $H_1:\theta<\theta_0$ and $H_1:\theta>\theta_0$ are called one-sided alternative. The alternative $H_1:\theta\neq\theta_0$ is a two-sided alternative

**p-value**: Let $T$ be a test statistic and $t_{obs}$ is the realized value of $T$, the **p-value** is defined by

$$\text{p-value}=P(T \text{ is as or more extreme than } t_{obs}|H_0)$$

A small p-value means the data are not consistent with $H_0$, and so is regarded as evidence against $H_0$

Suppose the observation $X=(X_1,\dots,X_n)$ have a joint distri $f(X|\theta)$. $H_0$ may specify that $\theta\in\omega_0$ where $\omega_0$ is a subset of the set of all possible values of $\theta$. Similarly, $H_1$ may specify $\theta\in\omega_1$. Let $\Omega=\omega_1\cup\omega_0$

If the hypotheses are composite, each likelihood is evaluated at that value of $\theta$ that maximizes it, yielding the **generalized likelihood ratio**

$$\Lambda=\max_{\theta\in\omega_0}[\text{lik}(\theta)]/\max_{\theta\in\Omega}[\text{lik}(\theta)]$$

The asymptotic null distribution of $-2\log(\Lambda)$ is $\chi^2_{r-r_0}$ distribution. where $r,r_0$ is the number of free parameters in $\Omega,\omega_0$, and we reject $H_0$ when $-2\log(\Lambda)>\chi^2_{r-r_0}(\alpha)$ where $\chi^2_n(\alpha)$ is the upper quantile

## Example

suppose $X$ represents a **single** observation from a population:

$$f(x\mid\theta)=\begin{cases}\frac{1}{\theta}2xe^{-x^2/\theta} & x>0\\ 0 & elsewhere\end{cases}$$

1. Show that the random variable $X^2$ has an exponential distribution with parameter $\theta$ 2. Find the uniformly most powerful test for testing $H_0:\theta=\theta_0, H_1:\theta>\theta_0$

consider a test $H_0:\theta=\theta_0, H_1^*:\theta=\theta_1$ and $\theta_1>\theta_0$, the likelihood ratio is

$$\Lambda(X)=\frac{\theta_1}{\theta_0}\exp\left\{-\sum X^2\left(\frac{1}{\theta_0}-\frac{1}{\theta_1}\right)\right\}$$

by L-P lemma we have the MP test is to reject $H_0$ when $\Lambda(X)<c$ for some $c$. since $\theta_1>\theta_0$, $\Lambda(X)$ is monotone decreasing w.r.t $X$, thus

$$\Lambda(X)<c\Leftrightarrow X^2>c_0$$

for some constant $c_0$, since $X^2\sim EXP(1/\theta)$, then

$$\alpha=P(X^2>c_0|H_0)=e^{-c_0/\theta_0}\implies c_0=-\theta_0\ln\alpha$$

thus the reject region is

$$R=\{X^2:X^2>-\theta_0\ln\alpha\}$$

which is free of $\theta_1$, thus the test with reject region $R$ is the UMP test for $H_0:\theta=\theta_0, H_1:\theta>\theta_0$

## Example

Let $X_1,\dots,X_n\sim EXP(1/\tau)$ with pdf

$$f(x|\tau)=1/\tau\exp\{-x/\tau\}$$

1. find MLE of $\tau$, $\hat\tau$
2. What is the exact sampling distribution of $\hat\tau$
3. show $\hat\tau$ is unbiased, and find its exact variance
4. Find an unbiased estimator $\tilde\tau$ based on $X_{(1)}$
5. distinguish which is better
6. Find the form of an approximate confidence interval for $\tau$
7. Find the form of an exact confidence interval for $\tau$
8. A theoretical model suggests that the time to breakdown of an insulating fluid between electrodes at a particular voltage has an exponential distribution with parameter $\lambda$, A random sample of $n=10$ breakdown times yields the following sample data 41.53, 18.73, 2.99, 30.34, 12.33, 117.52, 73.02, 223.63, 4.00, 26.78. We want to obtain an exact 95% confidence interval for $\lambda$ and the average breakdown time $\tau=1/\tau$
9. Find the most powerful test of $H_0:\tau=\tau_0$ vs $H_1:\tau=\tau_1$, where $\tau_1>\tau_0$. Give the rejection region EXPLICITLY in terms of a distribution function
10. Explain how you would compute the power of this test for $\tau=2$, assuming $\tau_0=1$ at level $\alpha=0.05$

Let $X_1,\dots,X_n\sim EXP(1/\tau)$ with pdf

$$L(\tau|x)=1/\tau^n\exp\left\{-1/\tau\sum x_i\right\}\implies l(\tau|x)=-n\ln\tau-\sum x_i/\tau$$

$$\implies -n/\tau+\sum x_i/\tau^2=0\implies\tau=\sum x_i/n\implies\hat\tau=\bar X$$

The mgf of $X_i$ is $M_{X_i}(t)=(1-\tau t)^{-1}$, then $M_{\sum X_i}(t)=(1-\tau t)^{-n}$,

$$M_{\hat\tau}(t)=M_{\sum X_i/n}(t)=M_{\sum X_i}(t/n)=\left(\frac{n/\tau}{n/\tau-t}\right)^n$$

then $\hat\tau\sim\Gamma(n,n/\tau)$

$$M'_{\hat\tau}(t)=n(n/\tau)^n(n/\tau-t)^{-n-1}\implies M'(0)=\tau$$

$$M''(t)=n(n+1)(n/\tau)^n(n/\tau-1)^{-n-2}\implies M''(0)=(n+1)\tau^2/n$$

$$\text{Var}[\hat\tau]=(n+1)\tau^2/n-\tau^2=\tau^2/n$$

consider

$$P(X_{(1)}>c)=\prod_{i=1}^n P(X_i>c)=\prod\exp\{-c/\tau\}=\exp\{-nc/\tau\}$$

$$\text{PDF is } F_{X_{(1)}}=1-\exp\{-nx/\tau\}\implies X_{(1)}\sim EXP(n/\tau)$$

thus $\mathbb{E}[X_{(1)}]=\tau/n$, so we let $\tilde\tau=nX(1)$, then $\mathbb{E}[\tilde\tau]=\tau$

the fisher information is

$$I(\tau)=-\mathbb{E}\left[\frac{n}{\tau^2}-\frac{2}{\tau^3}\sum X_i\right]=-\frac{n}{\tau^2}+\frac{2}{\tau^3}n\tau=\frac{n}{\tau^2}$$

thus the CR-lower bound is $1/I(\tau)=\tau^2/n$, which is equal to $\text{Var}[\hat\tau]$

By Asymptotic Normality

$$\sqrt{I(\tau)}(\hat\tau-\tau)\sim N(0,1)\implies\hat\tau\sim N\left(\tau,\frac{\tau^2}{n}\right)$$

thus $(1-\alpha)100\%$ asymptotic C.I. is $(\hat\tau\pm z(\alpha/2)\hat\tau/\sqrt{n})$

since $\sum X_i\sim\Gamma(n,1/\tau)$, let $Y=2\sum X_i/\tau\sim\Gamma(n,1/2)=\chi^2_{2n})$ not depend on $\tau$, using pivot $Y$, then

$$1-\alpha=P(\chi^2_{2n,1-\alpha/2}\le Y\le\chi^2_{2n,\alpha/2})$$

$$=P\left(\frac{2\sum X_i}{\chi^2_{2n,\alpha/2}}\le\tau\le\frac{2\sum X_i}{\chi^2_{2n,1-\alpha/2}}\right)$$

$\chi^2_{20,0.975}=9.59, \chi^2_{20,0.025}=34.17, \sum X_i=550.87$, the C.I. is $(32.24,114.87)$, for $\lambda=1/\tau$, the C.I. is $(1/114.87,1/32.24)=(0.0087,0.0310)$

$$\Lambda(X)=\frac{L(\tau_0)}{L(\tau_1)}=\left(\frac{\tau_1}{\tau_0}\right)^n\exp\left\{-\sum X_i\left(\frac{1}{\tau_0}-\frac{1}{\tau_1}\right)\right\}$$

since $\tau_1>\tau_0$, the $\Lambda(X)$ is a monotone decreasing funciton of $\sum X_i$, By N-P lemma, the MP test has reject region $R=\{X:\Lambda(X)<c_0\}=\{X:\sum X_i>c\}$. Under $H_0$, $2\sum X_i/\tau_0\sim\chi^2_{2n}$ , then

$$\alpha=P\left(2\sum X_i/\tau_0>\chi^2_{2n,\alpha}\right)$$

$$=P\left(\sum X_i>\tau_0\chi^2_{2n,\alpha}/2\right)\implies c=\frac{\tau_0\chi^2_{2n,\alpha}}{2}$$

$$\text{power}=P(\text{reject } H_1|H_1 \text{ is true})=P\left(\sum X_i<c|\tau=2\right)$$

$$=P\left(\sum X_i<\frac{1}{2}\chi^2_{2n,\alpha}|\tau=2\right)=P\left(\frac{2\sum X_i}{\tau}<\frac{\chi^2_{2n,\alpha}}{\tau}|\tau=2\right)$$

$$=F\left(\frac{1}{2}\chi^2_{2n,\alpha}\right)$$

where $F$ is the CDF of $\chi^2_{2n}$

## Example

Let $X_1,\dots,X_n$ be iid and $\mathbb{E}[X]=\sqrt{\theta\pi/2}, \text{Var}[X]=(2-\pi/2)\theta$

$$f(x|\theta)=\frac{x}{\theta}\exp\left\{-\frac{x^2}{\theta}\right\}, x>0$$

1. Find a sufficient statistic for $\theta$
2. Find the MLE $\hat\theta$ for $\theta$
3. Use the Cramer-Rao lower bound to justify if $\hat\theta$ is the minimum unbiased estimator of $\theta$
4. Specify asymptotic distribution of $\hat\theta$
5. Based on the asymptotic distribution of $\hat\theta$ obtained in part 4, find the form of an approximate $100(1-\alpha)\%$ confidence interval for $\theta$
6. Suppose from a random sample of 400 random variables you obtain $\sum_{i=1}^{400}x_i^2=3360$, Find the Likelihood Ratio test statistic for testing $H_0:\theta=4, H_1:\theta\neq 4$. specify the distribution and verify the null hypothesis at level $\alpha=0.05$

the joint pdf is

$$L(\theta|x)=\frac{\prod x_i}{\theta^n}\exp\left\{-\frac{\sum x_i^2}{2\theta}\right\}\to l(\theta|x)=-n\ln\theta-\sum\ln x_i-\frac{\sum x_i^2}{2\theta}$$

by factorization theorem, $g(T,\theta)=\theta^{-n}\exp\left\{-\sum x_i^2/2\theta\right\}, h(x)=\prod x_i$, then $\sum x_i^2$ is sufficient for $\theta$

$$\frac{\partial}{\partial\theta}l=-\frac{n}{\theta}+\sum\frac{x_i^2}{2\theta^2}=0\to\hat\theta=\frac{\sum x_i^2}{2n}$$

the Fisher information is

$$I(\theta)=-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}l\right]=-\frac{n}{\theta^2}+\sum\frac{\mathbb{E}[X_i^2]}{\theta^3}=-\frac{n}{\theta^2}+\frac{n}{\theta^3}\left(2\theta-\frac{\pi}{2}\theta+\frac{\theta\pi}{2}\right)=\frac{n}{\theta^2}$$

Let $Y=X^2$, then $X=\sqrt{Y}$, then

$$f_Y(y)=f_X(\sqrt{y})|\frac{d\sqrt{y}}{dy}|=\frac{1}{\theta}\sqrt{y}e^{-y/2\theta}\frac{1}{2\sqrt{y}}=\frac{1}{2\theta}e^{-y/2\theta}$$

which shows $Y\sim EXP(1/2\theta)$, then $\text{Var}[X^2]=\text{Var}[Y]=(2\theta)^2$, thus

$$\text{Var}[\hat\theta]=\text{Var}\left[\sum X_i^2/2n\right]=\sum\text{Var}[X_i^2]/4n^2=n4\theta^2/4n^2=\theta^2/n=1/I(\theta)$$

therefore $\hat\theta$ is the unbiased estimator and achieve CR lower bound

By Asymptotic Normality, $\sqrt{I(\theta)}(\hat\theta-\theta)=\sqrt{n}/\theta(\hat\theta-\theta)\sim N(0,1)$, thus $\hat\theta\sim N(\theta,\theta^2/n)$

a $(1-\alpha)100\%$ confidence interval can be obtained by

$$1-\alpha\approx P(-z(\alpha/2)<\sqrt{n}/\hat\theta(\hat\theta-\theta)<z(\alpha/2))$$

$$=P(\hat\theta-\hat\theta z(\alpha/2)/\sqrt{n}<\theta<\hat\theta+\hat\theta z(\alpha/2)/\sqrt{n})$$

where $z(\alpha/2)$ satisfics $\alpha/2=P(Z>z(\alpha/2)|Z\sim N(0,1)$

The likelyhood ratio is

$$\Lambda(X)=\frac{L(\theta|X)}{L(\hat\theta|X)}=\frac{\hat\theta^n}{\theta^n}\exp\left\{-\sum X_i^2\left(\frac{1}{2\theta}-\frac{1}{2\hat\theta}\right)\right\}$$

and $-2\ln\Lambda\sim\chi^2_1$, then

$$-2\ln\Lambda=2n\ln\frac{\theta}{\hat\theta}+\sum X_i\left(\frac{1}{\theta_0}-\frac{1}{\hat\theta}\right)$$

the decision rule is reject $H_0$ when $-2\ln\Lambda>\chi^2_{1,0.95}=3.84$, with $\hat\theta=\sum x_i^2/2n=3360/800=4.2$ and $\theta=4$, we have

$$-2\ln\Lambda=2\times 400\times\ln\frac{4}{4.2}+3360\left(\frac{1}{4}-\frac{1}{4.2}\right)=0.9676<3.84$$

thus we accept $H_0$

## Example

if $X_1\dots X_n$ are iid with pdf

$$f(x|\theta)=\theta x^{-2}, x>=\theta$$

a. Find a sufficient statistic for $\theta$. b. Find the MLE of $\theta$. c. Find the method of moments estimate of $\theta$.

$$f(\mathbf{x}\mid\theta)=\prod_{i=1}^n\theta x_i^{-2}I_{(x_i\ge\theta)}=\theta^n I_{(x_{(1)}\ge\theta)}\cdot\prod_{i=1}^n x_i^{-2}=g(T(\mathbf{x}),\theta)\cdot h(\mathbf{x})$$

where $T(X)=X_{(1)}$ is sufficient statistic

$$L(\theta\mid\vec{x})=\theta^n I_{(x_{(1)}\ge\theta)}\cdot(\prod_{i=1}^n x_i^{-2})=\begin{cases}\theta^n\cdot(\prod_{i=1}^n x_i^{-2}) & x_{(1)}\ge\theta\\ 0 & x_{(1)}<\theta\end{cases}$$

The maximum is achieved when $\theta\le x_{(1)}$, so MLE $\hat\theta=x_{(1)}$ $E(X)=\int_\theta^\infty x\theta x^{-2}dx=\int_\theta^\infty\theta x^{-1}dx=\theta\log x|_\theta^\infty=\infty$, thus MME does not exist