$E\{\sum c_i g_i(X)\} = \sum c_i E\{g_i(X)\}$; $Var(\sum c_i X_i) = \sum c_i^2 Var(X_i) + 2\sum_{i<j} c_i c_j Cov(X_i, X_j)$;

$\binom{x}{r} = \frac{x(x-1)\cdots(x-r+1)}{r!}$, $\binom{-x}{r} = (-1)^r\binom{x+r-1}{r}$, $n2^{n-1} = \sum_{i=1}^n i\binom{n}{i}$

$\binom{a+b}{a-r} = \sum_{i+j=a-r}\binom{a}{i}\binom{b}{j} = \sum_{k=0}^{a-r}\binom{a}{a-r-k}\binom{b}{k} = \sum_{k=0}^{a-r}\binom{a}{r+k}\binom{b}{k}$, $0 = \sum_{i=1}^n (-1)^i i\binom{n}{i}$ $\Leftarrow [(1\pm x)^n]'|_{x=1}$

$X \perp\!\!\!\perp Y \Leftrightarrow f(X) \perp\!\!\!\perp g(Y), \forall f(\cdot), g(\cdot)$.

$X \perp\!\!\!\perp Y$ if $f(x,y) = f(x)f(y)$ or $F(x,y) = F(x)F(y)$.

$S_{(X,Y)} = \{(x,y): f(x,y) > 0\}$ — support.

$I_S(z) = 1$ if $z \in S$, $I_S(z) = 0$ if $z \notin S$. $\mu_r' = E(X^r)$, $\mu_r = E(X-\mu)^r$

$\mu_r = \sum_{i=0}^r (-1)^i\binom{r}{i}\mu_{r-i}'\mu^i$, μ_3: a measure of asymmetry of f (pdf), $\frac{\mu_3}{\sigma^3}$: coefficient of skewness, μ_4: kurtosis (degree of flatness), $\frac{\mu_4}{\sigma^4}$: coefficient of kurtosis. (of f "near its center") (more peaked ~ than the normal density).

median: $Pr\{X \le med(X)\} = 0.5$ & $Pr\{X \ge med(X)\} = 0.5$

$F(\xi_q) = P(X \le \xi_q) \ge q$ (ξ_q min, discrete). $F(\xi_q) = P(X \le \xi_q) = q$ (continuous).

Theorem 1.3: $g: \mathbb{R} \to \mathbb{R}^+$, $Pr\{g(X) \ge c\} \le \frac{E\{g(X)\}}{c}$ $\forall c > 0$. $g(X) = (X-\mu)^2$, $c \to c^2\sigma^2$, $Pr\{|X-\mu| \ge c\sigma\} \le \frac{1}{c^2}$ Chebyshev

$g(X) = |X|^r$, $c \to c^r$, $Pr\{|X| \ge c\} \le \frac{E(|X|^r)}{c^r}$ → provided it exists. Markov. $g(x)$ convex $\Leftrightarrow$ $g''(x) \ge 0, \forall x \in S$.

Theorem 1.4: g convex and X in domain of g, $E(\frac{1}{X}) \ge \frac{1}{EX}$. $E(X^2) \ge (EX)^2 \Rightarrow Var(X) \ge 0$.

$$E\{g(X)\} \geqslant g(E(X))$$ p. both E ∃. (Jensen's Ineq)

Theorem 1.5: $\{E(XY)\}^2 \leqslant E(X^2)E(Y^2)$ with "=" iff $P(Y = cX) = 1$ ∃c. $[Cov(X,Y)]^2 \le Var(X)Var(Y)$. Pf: Let $h(t) = E(tX - Y)^2 \ge 0$, quadratic, $\Delta < 0$.

$E(g(X)) = E[E(g(X)|Y)]$ $= \int E(g(X)|Y=y) f_Y(y)\,dy$; $Var[g(X)] = E[Var(g(X)|Y)] + Var[E(g(X)|Y)]$; Pf: $E[Var(g(X)|Y)]$ = 展开, 用前一个.

$S_N = \sum_{i=1}^N X_i$, $E[S_N] = E(N)E(X)$, $Var(S_N) = E(N)Var(X) + Var(N)[E(X)]^2$.

$Cov(X_1, X_2) = E[(X_1-\mu_1)(X_2-\mu_2)] = E(X_1X_2) - EX_1 \cdot EX_2$. $\rho = Cor(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)\cdot Var(X_2)}}$

QUANTILE AND MEDIAN

The q-th quantile of a r.v. X with cdf $F(\cdot)$, denoted by ξ_q, is defined as the smallest real number ξ satisfying $F(\xi) = \Pr(X \leqslant \xi) \geqslant q$.

If X is continuous, then the q-th quantile of X is defined as the smallest real number ξ satisfying $F(\xi) = \Pr(X \leqslant \xi) = q$.

The 0.5-th quantile $\xi_{0.5}$ is defined as the median of X, denoted by $med(X)$.

Alternatively, the median of X satisfies

$$\Pr\{X \leqslant med(X)\} \geqslant 0.5 \quad \text{and} \quad \Pr\{X \geqslant med(X)\} \geqslant 0.5.$$

If X is a continuous r.v. with pdf $f(x)$, then the median of X satisfies

$$\int_{-\infty}^{med(X)} f(x)\,dx = 0.5 = \int_{med(X)}^{\infty} f(x)\,dx.$$

Bivariate Normal Distribution.

$N_d(\vec{x}|\vec{\mu}, \vec{\Sigma}) = \frac{1}{(\sqrt{2\pi})^d|\Sigma|^{\frac{1}{2}}} \times \exp\{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\vec{\Sigma}^{-1}(\vec{x}-\vec{\mu})\}$.

$\vec{\Sigma}$: pos. definite.

$(X_1, X_2)^T \sim N_2(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2, \rho)$

$E(X_1|X_2 = x_2) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$

$Var(X_1|X_2 = x_2) = \sigma_1^2(1-\rho^2)$.

$Cov(X_1, X_2) = \rho\sigma_1\sigma_2$

$Corr(X_1, X_2) = \rho$.

$X_1 \perp\!\!\!\perp X_2 \Leftrightarrow \rho = 0$.

(indep. ⇔ uncorr.)

$a_1X_1 + a_2X_2 \sim N(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2)$

Stochastic Representation

Eg. ① $X \sim Exp(\beta)$, $U \sim U(0,1)$ $\Rightarrow X \overset{d}{=} -\log(U)/\beta$

② $\{X_i\}_{i=1}^n \overset{i.i.d}{\sim} Exp(\beta)$, $Y \sim Gamma(n, \beta) \Rightarrow Y \overset{d}{=} \sum_{i=1}^n X_i$.

$f_X(x)$

$M_X(t) = E(e^{tX}) = E\{\sum_{n=0}^\infty \frac{(tX)^n}{n!}\} = \sum_{n=0}^\infty \frac{t^n}{n!}\mu_n'$. $\mu_n' = E(X^n) = \frac{d^nM_X(t)}{dt^n}\Big|_{t=0}$

$M_{a+bX}(t) = e^{at}M_X(bt)$, where a and b are two constants.

If $X \perp\!\!\!\perp Y$, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Jensen's inequality provides a lower bound: $M_X(t) \geqslant \exp\{tE(X)\}$.

Table 1.2 Discrete probability distributions

Distribution	pmf $p(x)$	Parameter	mgf $M_X(t)$	$E(X)$	$Var(X)$
Binomial	$\binom{n}{x}p^xq^{n-x}$, $x = 0,1,\ldots,n$	$0<p<1$	$(pe^t+q)^n$	np	npq
Poisson	$\lambda^x e^{-\lambda}/x!$, $x \in \mathbb{N}_0$	$\lambda > 0$	$\exp\{\lambda(e^t-1)\}$	λ	λ
Geometric	pq^{x-1}, $x \in \mathbb{N}_1$	$0<p<1$	$pe^t/(1-qe^t)$	$1/p$	q/p^2
Negative binomial	$\binom{x-1}{r-1}p^rq^{x-r}$, $x \in \mathbb{N}_r$	$0<p<1$	$\{pe^t/(1-qe^t)\}^r$	r/p	rq/p^2

NOTE: $q \doteq 1-p$ and $\mathbb{N}_r \doteq \{r, r+1, \ldots, \infty\}$.

Table 1.3 Continuous probability distributions

Distribution	pdf $f(x)$	Parameter	mgf $M_X(t)$	$E(X)$	$Var(X)$
Uniform	$1/(b-a)$, $a \leqslant x \leqslant b$	$a < b$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Beta	$Beta(x\|a,b)$, $0<x<1$	$a>0$, $b>0$	—	$\frac{a}{c}$	$\frac{ab}{c^2(c+1)}$
Exponential	$\lambda e^{-\lambda x}$, $x \geqslant 0$	$\lambda > 0$	$\lambda/(\lambda-t)$, $t<\lambda$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$Gamma(x\|\alpha,\beta)$, $x>0$	$\alpha>0$, $\beta>0$	$\{\beta/(\beta-t)\}^\alpha$, $t<\beta$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Normal	$N(x\|\mu,\sigma^2)$, $x \in \mathbb{R}$	$\mu \in \mathbb{R}$, $\sigma^2>0$	$\exp(\mu t + 0.5\sigma^2t^2)$	μ	σ^2

NOTE: $c \doteq a+b$.

① $$\Pr(X = x) = \left\{\sum_{y \in S_Y}\frac{\Pr(Y=y|X=x)}{\Pr(X=x|Y=y)}\right\}^{-1}$$

② $$\Pr(X=x) \propto \frac{\Pr(X=x|Y=y_0)}{\Pr(Y=y_0|X=x)}$$

IBF. requirements:

① $S_{(X,Y)} = S_X \times S_Y$. In general, $S_{(X,Y)} = S_{(X|Y)} \times S_Y = S_{(Y|X)} \times S_X$

② $(x,y) \in S_{(X,Y)}$.

for all $x \in S_X$ and an arbitrarily fixed $y_0 \in S_Y$.

$$\int_{S_Y} f_Y(y)\,dy = \int_{S_Y}\frac{f_{(Y|X)}(y|x)}{f_{(X|Y)}(x|y)}\cdot f_X(x)\,dy = f_X(x)\cdot\int_{S_Y}\frac{f_{(Y|X)}(y|x)}{f_{(X|Y)}(x|y)}\,dy,$$

we immediately have the following *point-wise formula*:

① $$f_X(x) = \left\{\int_{S_Y}\frac{f_{(Y|X)}(y|x)}{f_{(X|Y)}(x|y)}\,dy\right\}^{-1}, \quad \text{for any } x \in S_X. \tag{1.56}$$

② $$f_X(x) = \left\{\int_{S_X}\frac{f_{(X|Y)}(x|y_0)}{f_{(Y|X)}(y_0|x)}\,dx\right\}^{-1}\frac{f_{(X|Y)}(x|y_0)}{f_{(Y|X)}(y_0|x)}$$

for all $x \in S_X$ and an arbitrarily fixed $y_0 \in S_Y$.

③ $$f_X(x) \propto \frac{f_{(X|Y)}(x|y_0)}{f_{(Y|X)}(y_0|x)},$$

for all $x \in S_X$ and an arbitrarily fixed $y_0 \in S_Y$

p.d.f: 1. c.d.f ⇒ differentiate ⇒ p.d.f.

2. $g(y) = f(x) \times \left|\frac{dx}{dy}\right|$; $g(y) = \sum_{i=1}^{n} f(h_i^{-1}(y)) \times \left|\frac{d h_i^{-1}(y)}{dy}\right|$, $y = h(x)$.

$g(y_1, y_2) = f(x_1, x_2) \times \left|\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right|$ $\left|\frac{\partial(x_1, \cdots, x_n)}{\partial(y_1, \cdots, y_n)}\right|$

3. Moment generating func.

4. For any non-negative & measurable $g(\cdot)$.

• $E\{g(Y)\} = \int_{-\infty}^{\infty} g(h(x)) f_X(x)\,dx = \int g(y) f_Y(y)\,dy \Rightarrow f_Y(y)$ is p.d.f of $Y = h(X)$.

• $E\{g(Y)\} = E\{g(h(X_1, X_2))\} = \iint g(h(x_1, x_2)) f_{(X_1,X_2)}(x_1,x_2)\,dx_1dx_2 = \int g(y) f_Y(y)\,dy$ → p.d.f of Y.

• $E\{g(Y)\} = \sum_x g(h(x))\cdot P_X(x) = \sum_y g(y)\cdot P_Y(y)$ • $E\{g(h(X_1, X_2))\} = \sum_{x_1}\sum_{x_2} g(h(x_1,x_2))\cdot P_{(X_1,X_2)}(x_1,x_2) = \sum_y g(y)\cdot P_Y(y)$ → p.d.f of Y.

5. Mixture Technique.

• $f_X(x) = \int_{S_Y} f(x,y)\,dy = \int_{S_Y} f_Y(y) f_{(X|Y)}(x|y)\,dy$. • $f_X(x) = \sum_{k \in S_Y} P_k \cdot f_k(x)$

6. SR Technique.

Definition 2.1 (Function of random variables). Let $\{X_i\}_{i=1}^n \overset{iid}{\sim} F(x)$. An arbitrary function $T(X_1, \ldots, X_n)$ of $\{X_i\}_{i=1}^n$ is called a *statistic*.

$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ $E(\bar{X}) = \mu$, $Var(\bar{X}) = \sigma^2/n$

$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ $E(S^2) = \sigma^2$ — unbiased.

• If $\{X_i\}_{i=1}^n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$, $S^2 \perp\!\!\!\perp \bar{X}$ and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. (2.10)

⇒ S_1^2, S_2^2: sample variances of size n_1 & n_2 from $N(\mu_i, \sigma_i^2)$.

$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2S_1^2}{\sigma_1^2S_2^2} \sim F(n_1-1, n_2-1)$. — Theorem 2.4.

Let $G_r(x)$ denote the cdf of $X_{(r)}$, then

$$G_r(x) = \frac{1}{B(r, n-r+1)}\int_0^{F(x)} t^{r-1}(1-t)^{n-r}\,dt.$$

$$g_r(x) = \frac{d}{dx}G_r(x) = \frac{1}{B(r,n-r+1)}\cdot\frac{d}{dx}\int_0^{F(x)} t^{r-1}(1-t)^{n-r}\,dt = \frac{n!}{(r-1)!(n-r)!}f(x)F^{r-1}(x)\{1-F(x)\}^{n-r}.$$

$$G_n(x) = \Pr\{\max(X_1, \ldots, X_n) \leqslant x\} = \Pr(X_1 \leqslant x, \ldots, X_n \leqslant x) = F^n(x).$$

$$g_n(x) = \frac{dG_n(x)}{dx} = nf(x)F^{n-1}(x).$$

$$G_1(x) = \Pr(X_{(1)} \leqslant x) = 1 - \Pr\{\min(X_1, \ldots, X_n) > x\} = 1 - \Pr(X_1 > x, \ldots, X_n > x) = 1 - \{1 - F(x)\}^n.$$

$$g_1(x) = \frac{dG_1(x)}{dx} = nf(x)\{1-F(x)\}^{n-1}.$$

CLT: ① $\{X_n\}$ i.i.d. ② $\sigma^2 < +\infty$ $\Rightarrow \frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{d}{\to} Z$, as $n \to \infty$. $\dot\sim N(0,1)$. $Z \sim N(0,1)$.

Definition 2.2 (Convergence in distribution). Given a sequence of r.v.'s $\{X_n\}_{n=1}^\infty$. Let $F_n(x)$ be the cdf of X_n, if there exists a r.v. X with cdf $F(x)$ such that

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

for all points x at which $F(x)$ is continuous, then we say that $\{X_n\}_{n=1}^\infty$ converges *in distribution* or *in law* to X and write $X_n \overset{D}{\to} X$ or $X_n \overset{L}{\to} X$.

Theorem 2.5 (Equivalent result). Let $\{X_n\}_{n=1}^\infty$ be a sequence of r.v's. Assume that the mgf $M_{X_n}(t) = M(t;n)$ of X_n exists for $|t| < h$ for all n, and there exists a r.v. X with mgf $M(t)$ that exists for $|t| < h_1 < h$. If

$$\lim_{n\to\infty} M(t;n) = M(t), \Rightarrow X_n \overset{L}{\to} X.$$

It is possible that $\lim_{n\to\infty} F_n(x_0) \neq F(x_0)$ for such points x_0 at which $F(x)$ is discontinuous.

$X_n \overset{L}{\to} X \Longleftrightarrow$ as $n \to \infty$, $X_n \overset{d}{=} X$.

The procedure for proving $X_n \overset{L}{\to} X$ is as follows:

Step 1: Find $F_n(x)$.

Step 2: Find $F(x)$.

Step 3: Prove $F_n(x) \to F(x)$ as $n \to \infty$.

Definition 3.1 (A statistic). A function of one or more r.v.'s that does not depend on any unknown parameters is called a *statistic*.

Theorem 3.2 (Extension of Theorem 3.1). Let $\hat{\boldsymbol\theta}$ be the MLE of $\boldsymbol\theta = (\theta_1, \ldots, \theta_p)^\top \in \Theta$. If $\boldsymbol\eta_{r\times1} = \boldsymbol h(\boldsymbol\theta) = (h_1(\boldsymbol\theta), \ldots, h_r(\boldsymbol\theta))^\top$ for $1 \leqslant r \leqslant p$ is a many-to-few transformation between $\boldsymbol\theta$ and $\boldsymbol\eta$, then $\hat{\boldsymbol\eta} = \boldsymbol h(\hat{\boldsymbol\theta}) = (h_1(\hat{\boldsymbol\theta}), \ldots, h_r(\hat{\boldsymbol\theta}))^\top$ is the MLE of $\boldsymbol\eta$.

Score func: $S(\theta) = S(\theta; \mathbf{x}) \doteq \frac{d\ell(\theta)}{d\theta} = \ell'(\theta) = \frac{L'(\theta)}{L(\theta)}$

We call $I_n(\theta) = Var\{S(\theta)\} = Var_{\mathbf{x}}\{S(\theta; \mathbf{x})\}$ (3.18)

the *Fisher information*, which is a measure of the amount of information that $\mathbf{x}$ contains about the unknown parameter θ.

In many statistical problems, we have $E\{S(\theta)\} = 0$ so that (3.18) becomes

$$I_n(\theta) = E\{S^2(\theta; \mathbf{x})\} = E\left\{\left(\frac{d\log L(\theta;\mathbf{x})}{d\theta}\right)^2\right\} \tag{3.19}$$

If the support of the population density $f(x;\theta)$ does not depend on θ $\Rightarrow E\{S(\theta)\} = 0$.

Theorem 3.4 (Alternative expression). Let $I_n(\theta)$ denote the Fisher information. If $E\{S(\theta)\} = 0$, then

$$I_n(\theta) = E\left\{-\frac{d^2\log L(\theta;\mathbf{x})}{d\theta^2}\right\} = nI(\theta), \tag{3.23}$$

where

$$I(\theta) = E\left[\left\{\frac{d\log f(X;\theta)}{d\theta}\right\}^2\right] = E\left\{-\frac{d^2\log f(X;\theta)}{d\theta^2}\right\} \tag{3.24}$$

denotes the Fisher information for a single sample.

Theorem 3.3 (The general CR inequality). Let $\tau(\theta)$ be an arbitrary function of the unknown parameter θ. If (i) $\hat\theta = T(\mathbf{x})$ is an unbiased estimator of $\tau(\theta)$, and (ii) the support of the population density $f(x;\theta)$ does not depend on the parameter θ, then

$$Var(\hat\theta) \geqslant \frac{\{\tau'(\theta)\}^2}{I_n(\theta)}, \tag{3.20}$$

where $I_n(\theta)$ is the Fisher information. CR lower bound

If $E\{S(\theta, \vec{X})\} = 0$ and $Var\{S(\theta, \vec{X})\} = nI(\theta)$, then

1. $[nI(\theta)]^{1/2}(\hat\theta_n - \theta) \overset{L}{\to} Z \sim N(0,1)$ as $n \to \infty$.

2. $$\frac{\sqrt{nI(\theta)}[g(\hat\theta_n) - g(\theta)]}{g'(\theta)} \dot\sim N(0,1).$$

3. $\frac{S(\theta; \vec{X})}{\sqrt{nI(\theta)}} \overset{L}{\to} N(0,1)$ as $n \to \infty$. $\hat\theta_n \overset{P}{\to} \theta$ as $n \to \infty$

The MLE $\hat\theta_n$ is an asymptotically unbiased estimator of θ.

The MLE $\hat\theta_n$ is an asymptotically UMVUE because it reaches the CR lower bound in the sense that

$$\lim_{n\to\infty} \text{Eff}_{\hat\theta_n}(\theta) = \lim_{n\to\infty}\frac{1/I_n(\theta)}{Var(\hat\theta_n)} = 1.$$

The MLE $\hat\theta_n$ is asymptotically normally distributed.

Definition 4.1 (Pivotal quantity). Assume that $X_1, \ldots, X_n \overset{iid}{\sim} f(x;\theta)$ and $T = T(X_1, \ldots, X_n)$ is a sufficient statistic of θ. Let $P = P(T, \theta)$ be a function of T and θ. If the distribution of P does not depend on θ, then P is called a *pivotal quantity* or *pivot*.

E.g. $\{X_i\}_{i=1}^n \overset{i.i.d}{\sim} Exp(\theta) \Rightarrow 2\theta n\bar{X} \sim Gamma(\frac{2n}{2}, \frac{1}{2}) = \chi^2(2n)$.

When sample size is n; i.e., $X_1, \ldots, X_n \overset{iid}{\sim} f(x;\theta)$, from (4.2), we have

$$-2\sum_{i=1}^n \log F(X_i;\theta) \sim \chi^2(2n),$$

which is a pivotal quantity.

CI of Normal Mean:

1° known variance: $X_1, \cdots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma_0^2)$.

$P = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma_0} \sim N(0,1)$. $[\bar{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}]$

2° Unknown variance: $X_1, \cdots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$.

$P = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1)$. $[\bar{X} - t(\frac{\alpha}{2}, n-1)\frac{S}{\sqrt{n}}, \bar{X} + t(\frac{\alpha}{2}, n-1)\frac{S}{\sqrt{n}}]$.

$-z_{\frac{\alpha}{2}}$, 0, $z_{\frac{\alpha}{2}}$

CI of difference of two normal means:

1. known two variances: $\sigma_0 \triangleq (\sigma_1^2/n_1 + \sigma_2^2/n_2)^{\frac{1}{2}}$

$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_0} \sim N(0,1)$. $[\bar{X}_1 - \bar{X}_2 - z_{\frac{\alpha}{2}}\cdot\sigma_0, \bar{X}_1 - \bar{X}_2 + z_{\frac{\alpha}{2}}\cdot\sigma_0]$.

2. unknown but equal: $n' \triangleq n_1 + n_2 - 2$.

$S_p \triangleq \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n'}$. $\bar{X}_1 - \bar{X}_2 \mp t(\frac{\alpha}{2}, n'-1)\cdot S_p(\frac{1}{n_1} + \frac{1}{n_2})^{\frac{1}{2}}$

3. unknown two variances

$$T_{Welch} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \dot\sim t(\nu)$$

$\nu = \{\frac{C^2}{n_1-1} + \frac{(1-C)^2}{n_2-1}\}^{-1}$, $C = \frac{S_1^2/n_1}{S_1^2/n_1 + S_2^2/n_2}$.

The joint density of $X_{(r_1)}, \ldots, X_{(r_k)}$ $(1 \leqslant r_1 < \cdots < r_k \leqslant n;\ 1 \leqslant k \leqslant n)$ is, for $x_1 < \cdots < x_k$ (or $x_{(r_1)} < \cdots < x_{(r_k)}$),

$$g_{r_1\cdots r_k}(x_1, \ldots, x_k) = n!\left\{\prod_{i=1}^k f(x_i)\right\}\cdot\prod_{i=1}^{k+1}\frac{\{F(x_i) - F(x_{i-1})\}^{r_i - r_{i-1} - 1}}{(r_i - r_{i-1} - 1)!}, \tag{2.24}$$

where $x_0 = -\infty$, $x_{k+1} = +\infty$, $r_0 = 0$ and $r_{k+1} = n+1$.

The joint pdf of $X_{(r)}$ and $X_{(s)}$ $(1 \leqslant r < s \leqslant n)$ is, for $x < y$,

$$g_{rs}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}f(x)f(y) \times F^{r-1}(x)\{F(y) - F(x)\}^{s-r-1}\{1 - F(y)\}^{n-s}. \tag{2.25}$$

The joint pdf of $X_{(1)}, \ldots, X_{(r)}$ $(1 \leqslant r \leqslant n)$ is, for $x_1 < \cdots < x_r$,

$$g_{1\cdots r}(x_1, \ldots, x_r) = \frac{n!}{(n-r)!}f(x_1)\cdots f(x_r)\{1 - F(x_r)\}^{n-r}. \tag{2.26}$$

The joint pdf of $X_{(1)}, \ldots, X_{(n)}$ is, for $x_1 < \cdots < x_n$,

$$g_{1\cdots n}(x_1, \ldots, x_n) = n!f(x_1)\cdots f(x_n). \tag{2.27}$$

*Example of uncorrelated but dependent: Let $\phi(x)$ be the pdf of $N(0,1)$. Since $x^3\phi(x)$ is an odd function, we have $X \sim N(0,1)$;

$$E(XY) = E(X^3) = \int_{-\infty}^\infty x^3\phi(x)\,dx = 0.$$

$Y = X^2$, $Y \sim \chi^2(1)$. ⇒ dependent.

Note that $E(X) = 0$, we obtain

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)Var(Y)}} = 0.$$

In other words, X and Y are uncorrelated but surely dependent.

* Also ex. of $f(x,y)$ ∄! $S_{(X,Y)} = \{(x,y): x \in \mathbb{R}, y = x^2\}$ which is a curve and the *measure/area* of $\mathbb{S}_{(X,Y)}$ is zero.

mle: $\vec\theta^* = \arg\max_{\vec\theta \in \Theta} L(\vec\theta)$ or $\ell(\vec\theta)$.

MLE: $\hat{\vec\theta}$.

Moment Estimator: Use $EX (= \bar{X})$, EX^2, ⋯, to express parameter.

① calculate EX, ⋯ in form of $g(\vec\theta)$; ② express the parameter (解方程组).

Given a random sample $\mathbf{x} = (X_1, \ldots, X_n)^\top$, determine the joint density of $\mathbf{x}$ and $\boldsymbol\theta$:

Bayesian

$$f(\boldsymbol x, \boldsymbol\theta) = \text{Likelihood} \times \text{Prior} = f(\boldsymbol x|\boldsymbol\theta) \times \pi(\boldsymbol\theta) = \left\{\prod_{i=1}^n f(x_i|\boldsymbol\theta)\right\} \times \pi(\boldsymbol\theta), \tag{3.9}$$

where $\boldsymbol x = (x_1, \ldots, x_n)^\top$.

Determine the *posterior density* (i.e., the conditional density of $\boldsymbol\theta$ given $\mathbf{x} = \boldsymbol x$) of $\boldsymbol\theta$.

$$p(\boldsymbol\theta|\boldsymbol x) = \frac{f(\boldsymbol x, \boldsymbol\theta)}{f(\boldsymbol x)} = c^{-1}f(\boldsymbol x, \boldsymbol\theta) \propto f(\boldsymbol x, \boldsymbol\theta) = \text{Likelihood} \times \text{Prior}, \tag{3.10}$$

where $f(\boldsymbol x) = \int_\Theta f(\boldsymbol x, \boldsymbol\theta)\,d\boldsymbol\theta \doteq c$ is the normalizing constant of $p(\boldsymbol\theta|\boldsymbol x)$ because $\mathbf{x} = \boldsymbol x$ is given.

The Bayesian estimate of $\boldsymbol\theta$ (i.e., the conditional expectation of $\boldsymbol\theta$) is defined by

$$E(\boldsymbol\theta|\boldsymbol x) = \int_\Theta \boldsymbol\theta\cdot p(\boldsymbol\theta|\boldsymbol x)\,d\boldsymbol\theta. \tag{3.11}$$

bias: $b(\theta) = E\{\varphi(\vec{X})\} - \theta$.

Efficiency: $\text{Eff}_{\hat\theta}(\theta) \propto \frac{1}{Var(\hat\theta)}$.

Relative ~: $\frac{\text{Eff}_{\hat\theta_1}(\theta)}{\text{Eff}_{\hat\theta_2}(\theta)} = \frac{Var(\hat\theta_2)}{Var(\hat\theta_1)}$.

UMVUE: unbiased + smallest Var.

Efficient: Var = C-R lower bound

Efficiency of an unbiased estimator: $\text{Eff}_{\hat\theta}(\theta) = \frac{1/I_n(\theta)}{Var(\hat\theta)}$.

Prove completeness:

Assume a func. $h(T)$ satisfies $E\{h(T)\} = 0$, $\forall \theta \in \Theta$.

W.T.S: $Pr\{h(T) = 0\} = 1$.

Definition 3.7 (Sufficient statistic). A statistic $T(\mathbf{x})$ is said to be a *sufficient statistic* of θ if the conditional distribution of $\mathbf{x}$, given $T(\mathbf{x}) = t$, does not depend on θ for any value of t. In discrete case, this means that

证明不suff. $\Pr\{X_1 = x_1, \ldots, X_n = x_n; \theta|T(\mathbf{x}) = t\} = h(\boldsymbol x)$

Theorem 3.5 (Factorization theorem). A statistic $T(\mathbf{x})$ is a sufficient statistic of the unknown parameter θ iff the joint pdf (or pmf) can be written in the form

$f(\boldsymbol x; \theta) = g(T(\boldsymbol x), \theta) \times h(\boldsymbol x)$.

证明 suff. $$f(x_1, \ldots, x_n; \theta) = f(\boldsymbol x; \theta) = g(T(\boldsymbol x); \theta) \times h(\boldsymbol x), \tag{3.27}$$

where $h(\boldsymbol x)$ does not depend on θ, $g(T;\theta)$ is a function of both T and θ, and it depends on $x_1, \ldots, x_n$ only through T.

- For some problems, no single sufficient statistic exists.
- However, there will always exist joint sufficient statistics.

suff. v.s. joint suff.

Definition 3.8 (Joint sufficient statistics). Let $X_1, \ldots, X_n \overset{iid}{\sim} f(x;\boldsymbol\theta)$. The statistics $T_1(\mathbf{x}), \ldots, T_r(\mathbf{x})$ are said to be *jointly sufficient* if the conditional distribution of $\mathbf{x}$, given $T_1 = t_1, \ldots, T_r = t_r$, does not depend on $\boldsymbol\theta$.

Theorem 3.6 (Factorization theorem with joint sufficient statistics). A set of statistics $T_1(\mathbf{x}), \ldots, T_r(\mathbf{x})$ is jointly sufficient for the parameter vector $\boldsymbol\theta$ iff the joint pdf (or pmf) can be written in the form

$$f(x_1, \ldots, x_n; \boldsymbol\theta) = f(\boldsymbol x; \boldsymbol\theta) = g(T_1(\boldsymbol x), \ldots, T_r(\boldsymbol x); \boldsymbol\theta) \times h(\boldsymbol x), \tag{3.29}$$

where $h(\boldsymbol x)$ does not depend on $\boldsymbol\theta$, $g(T_1, \ldots, T_r; \boldsymbol\theta)$ depends on $x_1, \ldots, x_n$ only through $T_1, \ldots, T_r$.

Definition 3.9 (Completeness). Let $X_1, \ldots, X_n$ denote a random sample from the pdf (or pmf) $f(x;\theta)$ with parameter space Θ and let $T(\mathbf{x})$ be a statistic, where $\mathbf{x} = (X_1, \ldots, X_n)$. The statistic T is said to be *complete* if

$$E\{h(T)\} = 0 \quad \text{for all } \theta \in \Theta$$

implies that $h(T) = 0$ with probability 1; i.e.,

$$\Pr\{h(T) = 0\} = 1 \quad \text{for all } \theta \in \Theta,$$

where the function $h(T)$ is a statistic.

Theorem 3.7 (Lehmann–Scheffé theorem). Let $T(\mathbf{x})$ is a complete sufficient statistic for θ. If $g(T)$ is an unbiased estimator of $\tau(\theta)$, then $g(T)$ is the unique UMVUE for $\tau(\theta)$.

Shortest CI:

Ex. shortest CI of μ when σ^2 is known!

$X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma_0^2)$.

Sol: WLOG, let $\sigma_0 = 1$.

Let $\Phi(\cdot)$ be the c.d.f of $N(0,1)$ and $\phi(\cdot)$ be the pdf of $N(0,1)$.

$Z = \sqrt{n}(\bar{X} - \mu) \sim N(0,1)$

$1 - \alpha = Pr(\bar{X} + a \le \mu \le \bar{X} + b)$

$= Pr(\sqrt{n}a \le -Z \le \sqrt{n}b)$

$\Rightarrow \alpha_1 = Pr(-Z < \sqrt{n}a) = \Phi(\sqrt{n}a)$.

$\alpha - \alpha_1 = Pr(-Z > \sqrt{n}b) = \Phi(-\sqrt{n}b)$.

$\Rightarrow a = \Phi^{-1}(\alpha_1)/\sqrt{n}$.

$b = -\Phi^{-1}(\alpha - \alpha_1)/\sqrt{n}$.

width: $L(\alpha_1) = b - a$.

$\hat\alpha_1 = \arg\min_{0 \le \alpha_1 \le \alpha} L(\alpha_1) = \frac{\alpha}{2}$.

$x = \Phi^{-1}(y)$. $\frac{d\Phi^{-1}(y)}{dy} = \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{\phi(y)} = \frac{1}{\phi(\Phi^{-1}(y))}$.

⇒ equal-tail CI: shortest

sym. about $\bar{X}$ + $\phi(\cdot)$ sym about y-axis ⇒ equal height.

The CI of normal variance (§4.4):

– 27.1 If $\mu = \mu_0$ is known, use the pivotal quantity

$$\sum_{i=1}^n\frac{(X_i - \mu_0)^2}{\sigma^2} \sim \chi^2(n)$$

to construct a $100(1-\alpha)\%$ CI of σ^2 as follows:

$$\left[\frac{\sum_{i=1}^n(X_i - \mu_0)^2}{\chi^2(\alpha/2, n)}, \frac{\sum_{i=1}^n(X_i - \mu_0)^2}{\chi^2(1 - \alpha/2, n)}\right],$$

– 27.2 If μ is unknown, use the pivotal quantity

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

to construct a $100(1-\alpha)\%$ CI of σ^2 as follows:

$$\left[\frac{(n-1)S^2}{\chi^2(\alpha/2, n-1)}, \frac{(n-1)S^2}{\chi^2(1-\alpha/2, n-1)}\right], \tag{4.15}$$

$\alpha = \Pr\{Z > z_\alpha\}, \quad Z \sim N(0,1),$

$\alpha = \Pr\{t(n) > t(\alpha, n)\},$

$\alpha = \Pr\{\chi^2(n) > \chi^2(\alpha, n)\},$

$\alpha = \Pr\{F(n,m) > F(\alpha, n, m)\}.$

CI of the Ratio of two normal variances:

$\frac{\sigma_2^2S_1^2}{\sigma_1^2S_2^2} \sim F(n_1-1, n_2-1)$. $f(1-\frac{\alpha}{2}, \nu_1, \nu_2) = \frac{1}{f(\frac{\alpha}{2}, \nu_2, \nu_1)}$. $\frac{1}{F(\nu_1, \nu_2)} \overset{d}{=} F(\nu_2, \nu_1)$

$[\frac{S_1^2}{S_2^2}\cdot f(\frac{\alpha}{2}, \nu_2, \nu_1), \frac{S_1^2}{S_2^2}\cdot f(\frac{\alpha}{2}, \nu_2, \nu_1)]$.

Large sample CI: $1 - \alpha \approx Pr\{-z_{\frac{\alpha}{2}} \le \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma(\mu)} \le z_{\frac{\alpha}{2}}\}$ $= Pr\{L_1 \le \mu \le U_1\}$. $\dot\sim N(0,1)$

① $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. $\sigma^2 = \sigma^2(\mu)$. may use MLE $\hat\mu$!

② $E\{S(\theta; \vec{X})\} = 0$, $Var\{S(\theta; \vec{X})\} = nI(\theta)$.

$\Rightarrow -z_{\frac{\alpha}{2}} \le \frac{S(\theta; \vec{X})}{\sqrt{nI(\theta)}} \le z_{\frac{\alpha}{2}}$. $\dot\sim N(0,1)$.

③ $\hat\theta_n$: MLE of θ based on $\{X_i\}_{i=1}^n$.

$-z_{\frac{\alpha}{2}} \le \{nI(\theta)\}^{\frac{1}{2}}(\hat\theta_n - \theta) \le z_{\frac{\alpha}{2}}$ $\dot\sim N(0,1)$.

may use MLE $I(\hat\theta_n)$.

$\alpha(\theta) = \Pr(\text{Type I error}) = \Pr(\text{rejecting } H_0 | H_0 \text{ is true}) = \Pr(\mathbf{x} \in \mathbb{C} | \theta \in \Theta_0)$,

$\beta(\theta) = \Pr(\text{Type II error}) = \Pr(\text{accepting } H_0 | H_0 \text{ is false}) = \Pr(\mathbf{x} \in \mathbb{C}' | \theta \in \Theta_1)$,

where $\mathbb{C}$ is the critical region. where $\mathbb{C}'$ is the acceptance region.

$p(\theta) = \Pr(\text{rejecting } H_0 | \theta) = \Pr(\mathbf{x} \in \mathbb{C} | \theta)$ power func.

— If $\theta \in \Theta_0$,

$p(\theta) = \Pr(\mathbf{x} \in \mathbb{C} | \theta \in \Theta_0) = \alpha(\theta)$.

— If $\theta \in \Theta_1$,

$p(\theta) = \Pr(\mathbf{x} \in \mathbb{C} | \theta \in \Theta_1) = 1 - \beta(\theta)$.

Thm 5.1 $(N_1, \cdots, N_m)^\top \sim$ Multinomial $(n; p_1, \cdots, p_m)$, $n = \sum N_i$, $1 = \sum p_i$.

$Q_n = \sum_{j=1}^{m} \frac{(N_j - np_j)^2}{np_j} \xrightarrow{L} \chi^2(m-1)$ as $n \to \infty$

$N_j \sim \text{Bino}(n, p_j) \Rightarrow N_j \sim \text{Poisson}(np_j)$

$\frac{N_j - np_j}{\sqrt{np_j}} \sim N(0,1)$, $\left(\frac{N_j - np_j}{\sqrt{np_j}}\right)^2 \sim \chi^2(1)$

Definition 5.1 (Size of a test) Consider to test H_0: $\theta \in \Theta_0$ against H_1: $\theta \in \Theta_1 = \Theta - \Theta_0$. A test φ with critical region $\mathbb{C}$ is said to have size α if

$$\sup_{\theta \in \Theta_0} p_\varphi(\theta) = \sup_{\theta \in \Theta_0} \Pr(\mathbf{x} \in \mathbb{C} \mid \theta) = \sup_{\theta \in \Theta_0} \alpha_\varphi(\theta) = \alpha, \tag{5.8}$$

where $\mathbf{x} = (X_1, \ldots, X_n)^\top$.

Definition 5.2 (Most powerful test) A test φ with critical region $\mathbb{C}$ is said to be the *most powerful test* with size α for testing H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$, if $P_\varphi(\theta_0) = \alpha$ & $P_\varphi(\theta_1) \geq P_\chi(\theta_1)$ $\forall$ test χ with $P_\chi(\theta_0) \leq \alpha$.

Lemma 5.1 (Neyman–Pearson Lemma). Assume that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x; \theta)$. Let the likelihood function be $L(\theta) = L(\theta; \boldsymbol{x})$. Then a test φ with critical region

$$\mathbb{C} = \left\{ \boldsymbol{x} = (x_1, \ldots, x_n)^\top: \frac{L(\theta_0)}{L(\theta_1)} \leqslant k \right\} \tag{5.12}$$

and size α is the most powerful test of size α for testing H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$, where k is a value determined by the size α.

use $\alpha = \Pr(\boldsymbol{x} \in \mathbb{C} | \theta = \theta_0)$

The ratio

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$$

is referred to as a value of the LR statistic

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}^{\mathrm{R}})}{L(\hat{\theta})} \tag{5.23}$$

where $\hat{\theta}^{\mathrm{R}}$ denotes the restricted MLE of θ in Θ_0.

If $\Theta = \Theta^*$, then $\hat{\theta}$ in (5.23) is the unrestricted MLE of θ; if $\Theta \subset \Theta^*$, then $\hat{\theta}$ is the restricted MLE of θ in Θ.

Obviously, $0 < \lambda(\mathbf{x}) \leqslant 1$. $\mathbb{C} = \{\boldsymbol{x}: \lambda(\boldsymbol{x}) \leq \lambda_\alpha\}$, $\alpha = \sup_{\theta \in \Theta_0} \Pr\{\lambda(\boldsymbol{x}) \leq \lambda_\alpha | \theta\}$

If $P = P(T, \theta)$ is a pivot and H_0: $\theta = \theta_0$, then $P_0 = P(T, \theta_0)$ is the corresponding test statistic. H_0 ✓ $\Rightarrow$ $P_0 = P$.

Step 1: Find $\hat{\theta}^{\mathrm{R}}$, the restricted MLE of θ in Θ_0, and $\hat{\theta}$, the (un)restricted MLE of θ in Θ.

Step 2: Calculate the LR statistic

$$\lambda(\mathbf{x}) = L(\hat{\theta}^{\mathrm{R}})/L(\hat{\theta}). \tag{5.23}$$

Step 3: Determine λ_α or c in

$$\mathbb{C} = \{x: \lambda(x) \leqslant \lambda_\alpha\} = \{x: \log \lambda(x) \leqslant c\}$$

based on $\uparrow$ $(0,1)$.

$$\alpha = \Pr\{\log \lambda(x) \leqslant c | H_0 \text{ is true}\}$$

Determine C:

Step I. Express $\lambda(x)$ or $\log \lambda(x)$ as a function of a statistic Q, e.g., $\lambda(x) = h(Q)$, such that under H_0, Q follows a given distribution (e.g., Normal, Gamma, Chi-squares, t, F).

Step II. Prove that $h(Q)$ is concave with a maximum or convex with a minimum.

Step III. If $h(Q)$ is concave, then

$$h(Q) \leqslant c$$

is equivalent to

$$Q \leqslant c_1 \quad \text{or} \quad Q \geqslant c_2.$$

Step IV. If $h(Q)$ is convex, then $h(Q) \geqslant c$ is equivalent to

$$Q \leqslant c_1 \quad \text{or} \quad Q \geqslant c_2.$$

Step V. Use the equal-tail approach, let

$$\alpha/2 = \Pr(Q \leqslant c_1 | H_0)$$

and

$$\alpha/2 = \Pr(Q \geqslant c_2 | H_0)$$

to determine the c_1 and c_2.

Test on Normal Means:

① One-sample normal test (known var) $X_1, \cdots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma_0^2)$. H_0: $\mu = \mu_0$, H_1: $\neq$, $>$ or $<$.

CR approach: ① $Z = \frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \sim N(0,1)$

Test sta: $Z_0 \triangleq \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} = Z + \frac{\mu - \mu_0}{\sigma_0/\sqrt{n}} \overset{H_0 \checkmark}{=} Z$

② H_1: $\mu \neq \mu_0$ $\quad \mu > \mu_0$ $\quad \mu < \mu_0$

$\alpha = \Pr(|Z| \geq z_{\alpha/2})$ $\quad \Pr(Z \geq z_\alpha)$ $\quad \Pr(Z \leq -z_\alpha)$

$\mathbb{C} = \{\vec{x}: |z_0| \geq z_{\alpha/2}\}$ $\{\vec{x}: z_0 \geq z_\alpha\}$ $\{\vec{x}: z_0 \leq -z_\alpha\}$

numeral calculation: $z_0 \in \mathbb{C} \Rightarrow$ reject H_0

p-value approach:

p-v = $2\Pr(Z \geqslant |z_0|)$ (H_0 is true) if H_1: $\mu \neq \mu_0$,

= $\Pr(Z^2 \geqslant z_0^2)$ (H_0 is true) = $\Pr\{\chi^2(1) \geqslant z_0^2\}$,

p-v = $\Pr(Z \geqslant z_0)$ (H_0 is true) if H_1: $\mu > \mu_0$,

p-v = $\Pr(Z \leqslant z_0)$ (H_0 is true) if H_1: $\mu < \mu_0$,

② One-sample t test. H_0: $\mu = \mu_0$, H_1: $\neq$, $>$ or $<$.

$X_1, \cdots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, σ^2 known.

CR approach: ① $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$.

Test sta: $T_0 \triangleq \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = T + \frac{\mu - \mu_0}{S/\sqrt{n}}$.

② H_1: $\mu \neq \mu_0$ $\quad \mu > \mu_0$ $\quad \mu < \mu_0$

$\alpha = \Pr\{|T| \geq t(\frac{\alpha}{2}, n-1)\}$ $\quad T \geq t(\alpha, n-1)$ $\quad T \leq -t(\alpha, n-1)$

$\mathbb{C} = \{\vec{x}: |t_0| \geq t(\frac{\alpha}{2}, n-1)\}$ $\quad t_0 \geq t(\alpha, n-1)$ $\quad t_0 \leq -t(\alpha, n-1)$

p-value approach:

p-v = $2\Pr(T \geqslant |t_0|)$ (H_0 is true) if H_1: $\mu \neq \mu_0$

p-v = $\Pr(T \geqslant t_0)$ (H_0 is true) if H_1: $\mu > \mu_0$

p-v = $\Pr(T \leqslant t_0)$ (H_0 is true) if H_1: $\mu < \mu_0$

③ Two-sample t test.

$X_{i1}, \cdots, X_{in} \overset{\text{iid}}{\sim} N(\mu_i, \sigma^2)$, $i = 1, 2$.

H_0: $\mu_1 - \mu_2 = \delta$; H_1: $\neq$, $>$ or $<$.

CR approach: ① $n' = n_1 + n_2 - 1$. $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n' - 1}}$

$T^* \triangleq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{1/n_1 + 1/n_2}} \sim t(n'-1)$.

Test sta: $T_0^* \triangleq \frac{(\bar{X}_1 - \bar{X}_2) - \delta}{S_p \sqrt{1/n_1 + 1/n_2}}$

p-value approach:

② H_1: $\mu_1 - \mu_2 \neq \delta$ $\quad \mu_1 - \mu_2 > \delta$ $\quad \mu_1 - \mu_2 < \delta$

$\alpha = \Pr\{|T^*| \geq t(\frac{\alpha}{2}, n'-1)\}$ $\quad T^* \geq t(\alpha, n'-1)$ $\quad T^* \leq -t(\alpha, n'-1)$

$\mathbb{C} = \{\vec{x}: |t_0| \geq t(\frac{\alpha}{2}, n'-1)\}$ $\quad t_0 \geq t(\alpha, n'-1)$ $\quad t_0 \leq -t(\alpha, n'-1)$

p-v = $2\Pr(T^* \geqslant |t_0|)$ (H_0 is true) if H_1: $\mu_1 - \mu_2 \neq \delta$

p-v = $\Pr(T^* \geqslant t_0)$ (H_0 is true) if H_1: $\mu_1 - \mu_2 > \delta$

p-v = $\Pr(T^* \leqslant t_0)$ (H_0 is true) if H_1: $\mu_1 - \mu_2 < \delta$

When p-value $\geqslant \alpha$ (level of significance), we cannot reject the H_0

χ^2 test for totally known distribution: Sample space $S = \bigcup_{j=1}^m A_j$. $X_1, \cdots, X_n \overset{\text{iid}}{\sim} F(x; \theta)$

$N_j = \#$ of $X_1, \cdots, X_n$ that fall in A_j. $(N_1, \cdots, N_m)^\top \sim$ Multinomial$(n; p_1, \cdots, p_m)$

$p_j = \Pr(X \in A_j) = \int_{A_j} d F(x; \theta)$. estimate p_j by $\hat{p}_j = \frac{N_j}{n}$, $n = \sum_j^m N_j$.

H_0: $p_j = p_{j0}$, $\forall j = 1, \cdots, m-1$. H_1: $p_j \neq p_{j0}$ for at least one of $j = 1, \cdots, m-1$.

$p_{j0} = \int_{A_j} d F_0(x; \theta)$. $Q_{n0} = \sum_{j=1}^{m} \frac{(N_j - np_{j0})^2}{np_{j0}} \overset{H_0}{=} Q_n \xrightarrow{L} \chi^2(m-1)$ as $n \to \infty$

If $Q_{n0} > \chi^2(\alpha, m-1)$, H_0 is rejected at the α significance level, when H_0 is true.

χ^2 test for known distribution family with unknown parameters: Let the MLE of θ_k be $\hat{\theta}_k$, $k = 1, \cdots, q$. $\Rightarrow$ MLE of p_{j0} is $\hat{p}_{j0} = p_{j0}(\hat{\theta}_1, \cdots, \hat{\theta}_q)$, $j = 1, \cdots, m$.

$p_{j0} = \int_{A_j} d F_0(x; \theta) = p_{j0}(\theta) = p_{j0}(\theta_1, \cdots, \theta_q)$

$\hat{Q}_{n0} = \sum_{j=1}^{m} \frac{(N_j - n\hat{p}_{j0})^2}{n\hat{p}_{j0}} = \hat{Q}_n \xrightarrow{L} \chi^2(m-q-1)$ as $n \to \infty$

If $\hat{Q}_{n0} > \chi^2(\alpha, m-q-1)$, H_0 is rejected at the α significance level.

A.1.1 Finite discrete distribution

Notation: $X \sim \text{FDiscrete}_n(\boldsymbol{x}, \boldsymbol{p})$, $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$, $\boldsymbol{p} = (p_1, \ldots, p_n)^\top \in \mathbb{T}_n \hat{=} \{(p_1, \ldots, p_n): p_i > 0, \sum_{i=1}^n p_i = 1\}$.

Density: $\Pr(X = x_i) = p_i, \quad i = 1, \ldots, n$.

Moments: $E(X) = \sum_{i=1}^n x_i p_i$ and $\text{Var}(X) = \sum_{i=1}^n x_i^2 p_i - (\sum_{i=1}^n x_i p_i)^2$.

Note: The *uniform discrete* distribution is a special case of the finite discrete distribution with $p_i = 1/n$ for all i.

$\Rightarrow X \sim$ Bernoulli(p). $\begin{array}{c|cc} X & 0 & 1 \\ \hline P_r & 1-p & p \end{array}$ $\Pr(X=x) = p^x(1-p)^{1-x}$. $EX = p$. $EX^n = p$. $\text{Var}X = p - p^2$. $X^n \overset{d}{=} X$, $n \in \mathbb{N}^*$.

$X = \frac{Y - x_1}{x_2 - x_1}$ gives 1-to-1 mapping 2 pt. distribution: $\begin{array}{c|cc} Y & x_1 & x_2 \\ \hline P_r & 1-p & p \end{array}$ $X = \frac{Y - x_1}{x_2 - x_1} = \begin{cases} 0, & Y = x_1 \\ 1, & Y = x_2 \end{cases}$ $Y \overset{d}{=} x_1 + (x_2 - x_1)X$

$X(1-X) \sim$ Degenerate (0).

A.1.2 Hypergeometric distribution

Notation: $X \sim \text{Hgeometric}(m, n, k)$, m, n, k are positive integers.

Density: $\text{Hgeometric}(x|m, n, k) = \binom{m}{x}\binom{n}{k-x}/\binom{m+n}{k}$, where $x = \max(0, k-n), \ldots, \min(m, k)$.

Moments: $E(X) = km/N'$ and $\text{Var}(X) = kmn(N' - k)/[N'^2(N' - 1)]$, where $N' \hat{=} m + n$.

A.1.3 Poisson distribution

Notation: $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$

Density: $\text{Poisson}(x|\lambda) = \lambda^x \mathrm{e}^{-\lambda}/x!$, $x = 0, 1, \ldots, \infty$.

Moments: $E(X) = \lambda$ and $\text{Var}(X) = \lambda$.

Properties: • If $\{X_i\}_{i=1}^n \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_i)$, then

$$\sum_{i=1}^n X_i \sim \text{Poisson}(\textstyle\sum_{i=1}^n \lambda_i), \text{ and}$$
$$(X_1, \ldots, X_n) | (\textstyle\sum_{i=1}^n X_i = m) \sim \text{Multinomial}_n(m, \boldsymbol{p}),$$

where $\boldsymbol{p} = (\lambda_1, \ldots, \lambda_n)^\top / \sum_{i=1}^n \lambda_i$;

• The Poisson and gamma distribution have relationship:

$$\textstyle\sum_{x=k}^\infty \text{Poisson}(x|\lambda) = \int_0^\lambda \text{Gamma}(y|k, 1)\, \mathrm{d}y.$$

$\{X_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma^2)$, $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$

$\Rightarrow T = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1)$.

A.1.4 Binomial distribution

Notation: $X \sim \text{Binomial}(n, p)$, n is a positive integer, $p \in (0, 1)$.

Density: $\text{Binomial}(x|n, p) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$.

Moments: $E(X) = np$ and $\text{Var}(X) = np(1-p)$.

Properties: • If $\{X_i\}_{i=1}^d \overset{\text{ind}}{\sim} \text{Binomial}(n_i, p)$, then

$$\textstyle\sum_{i=1}^d X_i \sim \text{Binomial}(\sum_{i=1}^d n_i, p);$$

• The binomial and beta distribution have relationship:

$$\textstyle\sum_{x=0}^k \text{Binomial}(x|n, p) = \int_0^{1-p} \text{Beta}(x|n-k, k+1)\, \mathrm{d}x,$$

where $0 \leqslant k \leqslant n$.

Note: When $n = 1$, binomial distribution is called *Bernoulli* distribution.

If $\{X_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma^2)$, $S^2 \perp\!\!\!\perp \bar{X}$ and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

A.1.5 Multinomial distribution

Notation: $\boldsymbol{x} = (X_1, \ldots, X_d)^\top \sim \text{Multinomial}(n; p_1, \ldots, p_d)$ or $\boldsymbol{x} = (X_1, \ldots, X_d)^\top \sim \text{Multinomial}_d(n, \boldsymbol{p})$, n is a positive integer, $\boldsymbol{p} = (p_1, \ldots, p_d)^\top \in \mathbb{T}_d$,

Density: $\text{Multinomial}_d(\boldsymbol{x}|n, \boldsymbol{p}) = \binom{n}{x_1, \ldots, x_d} \prod_{i=1}^d p_i^{x_i}$, $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$, $x_i \geqslant 0$, $\sum_{i=1}^d x_i = n$.

Moments: $E(X_i) = np_i$, $\text{Var}(X_i) = np_i(1-p_i)$ and $\text{Cov}(X_i, X_j) = -np_ip_j$.

Note: The binomial distribution is a special case of the multinomial with $d = 2$.

A.2.1 Uniform distribution

Notation: $X \sim U(a, b)$ or $X \sim U[a, b]$, $a < b$

Density: $U(x|a, b) = 1/(b-a)$, $x \in (a, b)$ or $x \in [a, b]$.

Moments: $E(X) = (a+b)/2$ and $\text{Var}(X) = (b-a)^2/12$.

Properties: If $Y \sim U(0, 1)$, then $X = a + (b-a)Y \sim U(a, b)$.

A.2.2 Beta distribution

$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $\int_0^1 x^{a-1}(1-x)^{b-1} dx = B(a, b)$

$S_x = \begin{cases} [0,1], & \text{if } a = 1, b = 1 \\ (0,1], & \text{if } a = 1, b \neq 1 \\ [0,1), & \text{if } a \neq 1, b = 1 \\ (0,1), & \text{if } a \neq 1, b \neq 1 \end{cases}$

Notation: $X \sim \text{Beta}(a, b)$, $a > 0, b > 0$.

Density: $\text{Beta}(x|a, b) = x^{a-1}(1-x)^{b-1}/B(a, b)$, $0 < x < 1$.

Moments: $E(X) = a/(a+b)$, $E(X^2) = a(a+1)/[(a+b)(a+b+1)]$ and $\text{Var}(X) = ab/[(a+b)^2(a+b+1)]$.

Properties: If $Y_1 \sim \text{Gamma}(a, 1)$, $Y_2 \sim \text{Gamma}(b, 1)$, and $Y_1 \perp\!\!\!\perp Y_2$, then $Y_1/(Y_1 + Y_2) \sim \text{Beta}(a, b)$.

$E(X^r) = \frac{B(a+r, b)}{B(a, b)} = \frac{\Gamma(a+r)}{\Gamma(a)} \cdot \frac{\Gamma(a+b)}{\Gamma(a+b+r)}$.

Note: When $a = b = 1$, $\text{Beta}(1, 1) = U(0, 1)$.

A.2.3 Exponential distribution

Notation: $X \sim \text{Exponential}(\beta)$, rate parameter $\beta > 0$.

Density: $\text{Exponential}(x|\beta) = \beta\,\mathrm{e}^{-\beta x}$, $x \geqslant 0$.

Moments: $E(X) = 1/\beta$ and $\text{Var}(X) = 1/\beta^2$.

Properties: • If $U \sim U(0, 1)$, then $-\log(U)/\beta \sim \text{Exponential}(\beta)$;

• If $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} \text{Exponential}(\beta)$, then $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \beta)$.

A.2.4 Gamma distribution

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, $\Gamma(1) = 1$; $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$; $\Gamma(p)\Gamma(1-p) = \frac{\pi}{\sin p\pi}$; $\Gamma(n+1) = n!$

Notation: $X \sim \text{Gamma}(\alpha, \beta)$, shape parameter $\alpha > 0$, rate parameter $\beta > 0$. $\int_0^\infty x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$

$S_x = \begin{cases} (0, +\infty), & \alpha \neq 1 \\ [0, +\infty), & \alpha = 1 \end{cases}$

Density: $\text{Gamma}(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}$, $x > 0$.

Moments: $E(X) = \alpha/\beta$ and $\text{Var}(X) = \alpha/\beta^2$. $E[\log(X)] = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \log(\beta)$. Exp$(\beta)$.

Properties: • If $X \sim \text{Gamma}(\alpha, \beta)$ and $c > 0$, then $cX \sim \text{Gamma}(\alpha, \beta/c)$;

• If $\{X_i\}_{i=1}^n \overset{\text{ind}}{\sim} \text{Gamma}(\alpha_i, \beta)$, then $\sum X_i \sim \text{Gamma}(\sum \alpha_i, \beta)$;

• $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$.

Note: $\text{Gamma}(1, \beta) = \text{Exponential}(\beta)$. $\text{Gamma}(\nu/2, 1/2) = \chi^2(\nu)$.

A.2.5 Chi-square distribution

Notation: $X \sim \chi^2(\nu) \equiv \text{Gamma}(\frac{\nu}{2}, \frac{1}{2})$, degrees of freedom $\nu > 0$.

Density: $\chi^2(x|\nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} \mathrm{e}^{-x/2}$, $x > 0$.

Moments: $E(X) = \nu$ and $\text{Var}(X) = 2\nu$.

Properties: • If $Y \sim N(0, 1)$, then $X = Y^2 \sim \chi^2(1)$;

• If $\{X_j\}_{j=1}^m \overset{\text{ind}}{\sim} \chi^2(\nu_j)$, then $\sum_{j=1}^m X_j \sim \chi^2(\sum_{j=1}^m \nu_j)$.

A.2.6 *t*- or Student's *t*-distribution

Notation: $X \sim t(\nu)$, $\nu > 0$ is a positive real number.

Density: $t(x|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < x < \infty$.

Moments: $E(X) = 0$ (if $\nu > 1$) and $\text{Var}(X) = \frac{\nu}{\nu-2}$ (if $\nu > 2$).

Properties: Let $Z \sim N(0, 1)$, $Y \sim \chi^2(\nu)$, and $Z \perp\!\!\!\perp Y$, then

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1). \qquad \frac{Z}{\sqrt{Y/\nu}} \sim t(\nu).$$

A.2.7 *F* or Fisher's *F*-distribution

Notation: $X \sim F(n_1, n_2)$, n_1, n_2 are positive integers.

Density: $F(x|n_1, n_2) = \frac{(n_1/n_2)^{n_1/2}}{B(\frac{n_1}{2}, \frac{n_2}{2})} x^{\frac{n_1}{2}-1}(1 + \frac{n_1 x}{n_2})^{-\frac{n_1+n_2}{2}}$, $x > 0$.

Moments: $E(X) = \frac{n_2}{n_2-2}$ (if $n_2 > 2$), $\text{Var}(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)(n_2-2)^2}$ (if $n_2 > 4$).

Properties: Let $Y_i \sim \chi^2(n_i)$, $i = 1, 2$, and $Y_1 \perp\!\!\!\perp Y_2$, then

$$\frac{Y_1/n_1}{Y_2/n_2} \sim F(n_1, n_2).$$

A.2.8 Normal or Gaussian distribution

Notation: $X \sim N(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$.

Density: $N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$, $-\infty < x < \infty$.

Moments: $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. $a + bX \sim N(a + b\mu, b^2\mu^2)$

Properties: • If $\{X_i\} \overset{\text{ind}}{\sim} N(\mu_i, \sigma_i^2)$, then $\sum a_i X_i \sim N(\sum a_i\mu_i, \sum a_i^2\sigma_i^2)$;

• If $X_1 | X_2 \sim N(X_2, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, then $X_1 \sim N(\mu_2, \sigma_1^2 + \sigma_2^2)$.

A.2.9 Multivariate normal or Gaussian distribution

Notation: $\mathbf{x} = (X_1, \ldots, X_d)^\top \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ or $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} > 0$.

Density: $N_d(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(\sqrt{2\pi})^d |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\}$, $\boldsymbol{x} \in \mathbb{R}^d$.

Moments: $E(\mathbf{x}) = \boldsymbol{\mu}$ and $\text{Var}(\mathbf{x}) = \boldsymbol{\Sigma}$.

Multivariate Normal Distribution

$\begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} + A \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} \Rightarrow \vec{X} \sim N_d(\vec{\mu}, AA^\top)$, $Y_i \sim N(0,1)$.

Advantage: (Note: $AA^\top$ pos semi-def.) ① If $AA^\top$ pos def. $\Rightarrow \exists (AA^\top)^{-1} \Rightarrow$ joint density of $\vec{X}$ exists, If $|AA^\top| = 0$, the joint density $\nexists$ but distribution $\exists$. ② Easier to see: $E(\vec{X}) = \vec{\mu}$, $\text{Var}(\vec{X}) = \text{Cov}(\vec{X}, \vec{X}) = AA^\top$. $M_{\vec{X}}(\vec{t}) = E(e^{\vec{t}^\top \vec{X}}) = \exp(\vec{t}^\top \vec{\mu} + \frac{1}{2}\vec{t}^\top AA^\top \vec{t})$.

指数分布的密度函数、累积分布函数和生存函数

设 $X \sim \text{Exponential}(\theta)$, 则它的 pdf, cdf 和生存函数分别定义为

$f(x) = \theta\exp(-\theta x), \quad x \geqslant 0, \quad \theta > 0,$

$F(x) = \Pr(X \leqslant x) = \int_0^x f(t)\,\mathrm{d}t = 1 - \exp(-\theta x),$

$S(x) = \Pr(X > x) = 1 - F(x) = \exp(-\theta x).$

指数分布的无记忆性

对指数随机变量 X(例如它是某一元件的寿命), 当 $s, t > 0$, 我们有

$$\Pr(X > t + s | X > t) = \Pr(X > s) = \exp(-\theta s). \tag{11.1}$$

注释: 在所有的连续分布中, 指数分布是唯一的具有无记忆性的连续分布。

几何分布的概率密度函数和生存函数分别

设 $X \sim \text{Geometric}(p)$, 则它的 pmf 和生存函数分别定义为

$\Pr(X = x) = (1-p)^x p, \quad x = 0, 1, \ldots, \infty; \quad p \in (0, 1),$

$\Pr(X > m) = \sum_{x=m+1}^{\infty} \Pr(X = x) = (1-p)^{m+1}.$

几何分布的无记忆性

对几何随机变量 X, 当 n, m 为两个正整数, 我们有

$$\Pr(X > n + m | X > n) = \Pr(X > m) = (1-p)^{m+1}. \tag{11.2}$$

注释: 在所有的离散分布中, 几何分布是唯一的具有无记忆性的离散分布。

6° 几何分布是指数分布的离散影像

• 设 $Y \sim \text{Exponential}(\theta)$ with $\theta > 0$, 且 $X = \lfloor Y \rfloor$, 我们有

$$\Pr(X = x) = \exp[-\theta(x+1)] - \exp(-\theta x) = (1-p)^x p, \tag{11.4}$$

其中 $x = 0, 1, \ldots, \infty$, $p \triangleq 1 - \mathrm{e}^{-\theta} = F_Y(1) \in (0, 1)$。

二项分布的泊松近似

设 $\{Z_n\}_{n=1}^\infty$ 是随机变量序列, $Z_n \sim \text{Binomial}(n, p)$ 且 $np = \mu$, 则

$$Z_n \xrightarrow{L} \text{Poisson}(\mu), \tag{14}$$

as $n \to \infty$。见图 14.1。正态近似: $\frac{Z_n - np}{\sqrt{np(1-p)}} \xrightarrow{L} N(0,1)$.

3° Monte Carlo 积分公式

• 设 $X_1, \ldots, X_n, X \overset{\text{iid}}{\sim} g(x; \theta)$, 其中 $g(x; \theta)$ 是一个密度函数。对任意的一个函数 $h(X)$, 假设它的定义域 $\mathbb{X}$ 与 $g(x; \theta)$ 的定义域相同, 则

$$E\{h(X)\} = \int_{\mathbb{X}} h(x) \cdot g(x; \theta)\,\mathrm{d}x = \frac{1}{n}\sum_{i=1}^n h(X_i), \quad n \to \infty, \tag{15.2}$$

称为 Monte Carlo 积分公式, 它是 (15.1) 的推广。

4° 证明公式 (15.2)

• 定义 $Y_i = h(X_i)$, $i = 1, \ldots, n$, 且 $Y = h(X)$, 当 $n \to \infty$, 则

$$\frac{1}{n}\sum_{i=1}^n h(X_i) = \frac{1}{n}\sum_{i=1}^n Y_i \overset{(15.1)}{=} E(Y) = E\{h(X)\}.$$

• Let $h(x) = 0$, we obtain the $(t+1)$-th approximation of $x^{(\infty)}$ as

want: roots of $g(x)$.

$$x^{(t+1)} = x^{(t)} - \frac{g(x^{(t)})}{g'(x^{(t)})}, \quad t = 0, 1, 2, \ldots \tag{C.2}$$

• By comparing (C.4) with (C.1); i.e., by replacing $g(\cdot)$ in (C.2) with $\ell'(\cdot)$, we obtain

want: MLE $\hat{\theta}$ $\Leftrightarrow$ roots of $\ell'(\theta)$. $\ell(\theta)$ — log likelihood func.

$$\theta^{(t+1)} = \theta^{(t)} - \frac{\ell'(\theta^{(t)})}{\ell''(\theta^{(t)})}, \quad t = 0, 1, 2, \ldots, \tag{C.5}$$

which defines Newton's method for finding MLE.