ST2132 Mathematical Statistics

Lingjie, May 8, 2021

Flow of analysis

1. Proposal → Identify distribution

   [Discrete] i.i.d.

   [Discrete] multinomial

   [Continuous] i.i.d.

2. Data → sufficient statistics

3. Training → finding parameters $\hat{\theta}$

   [method] MoM, MLE

   [kind] point estimate, confidence interval

4. Compare performance → varaince, bias trade off

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

5. Evaluate → goodness of fit

   [general] likelihood ratio test

   [discrete] perason chi-sq statistics

6. A/B testing → comparing average effect

   two sample mean test

Review of Probability

Conditional Probability

Definition 1: conditional probaility

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Theorem 1: Law of Total Probability & Bayes' Rule

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$
$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

Independent

Definition 2: independent event

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

Pairwise independence does not guarantee mutal independence. Mutal independence:

$$P(A_{i1} \cap \cdots \cap A_{im}) = P(A_{i1}) \cdots P(A_{im})$$

Definition 3: independent RV

$$F(X_1, x_2, \cdots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

Functions of a RV

Proposition 1

$X \sim N(\mu, \sigma^2), Y = aX + b \Rightarrow Y \sim N(\alpha\mu + b, a^2\sigma^2)$

Proposition 2

$Y = g(X) \Rightarrow f_Y(y) = f_X\left[g^{-1}(y)\right] |\frac{d}{dy}g^{-1}(y)|$

Note:

When function is not strictly monotonic (e.g. $g(z) = z^2$), proposition 2 cannot be used. Instead, solve $F_x(x) = P(X \leq x) = P(Z^2 \leq x) = P(-\sqrt{x} \leq Z \leq \sqrt{x}) = F_Z(\sqrt{x}) - F_Z(-\sqrt{x}), x \geq 0$

Multinomial Distribution

$n :=$ num of independent trials

$r :=$ num of types

$X_i :=$ total number of outcomes of type i in the n trials

$$p(x_1, x_2, \cdots, x_r) = \binom{n}{x_1 \cdots x_r} p_1^{x_1}, p_2^{x_2} \cdots p_r^{x_r}$$

note: multi-nomial are not independent

Quotient of two continuous RV

Given $f(x, y)$ and $Z = Y/X$ then

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(\frac{Y}{X} \leq z) \\ &= P(X \leq 0, Y \geq Xz) + P(X > 0, Y \leq Xz) \\ &= \int_{-\infty}^{0}\int_{xz}^{\infty} f(x,y)dydx + \int_{0}^{\infty}\int_{-\infty}^{xz} f(x,y)dydx \end{aligned}$$

let $v := y/x$

$$\begin{aligned} &= \int_{-\infty}^{0}\int_{-\infty}^{z} (-x)f(x,xv)dvdx + \int_{0}^{\infty}\int_{-\infty}^{z} xf(x,xv)dvdx \\ &= \int_{-\infty}^{z}\int_{-\infty}^{\infty} |x|f(x,xv)dxdv \end{aligned}$$

$\therefore f_Z(z) = \int_{-\infty}^{\infty} |x|f(x, xz)dx$

if X, Y independent $\Rightarrow f_Z(z) = \int_{-\infty}^{\infty} |x|f_X(x)f_Y(xz)dx$

Extrema

$X_1, X_2, \cdots, X_n$ are i.i.d RV with F, f

Maximum: $U = \max\{X_1, X_2, \cdots, X_n\}$

For given u $U \leq u \Leftrightarrow X_i \leq u$

$$\begin{aligned} F_U(u) &= P(U \leq u) \\ &= P(X_1 \leq u) \cdots P(X_n \leq u) \\ &= F(u)^n \\ f_U(u) &= nf(u)F(u)^{n-1} \end{aligned}$$

Minimum: $V = \min\{X_1, X_2, \cdots, X_n\}$

For given v, $V \geq u \Leftrightarrow X_i \geq v$

$$\begin{aligned} 1 - F_V(v) &= P(V \geq v) \\ &= P(X_1 \geq v) \cdots P(X_n \geq v) \\ &= [1 - F(v)]^n \\ \Rightarrow F_V(v) &= 1 - [1 - F(v)]^n \\ f_V(v) &= nf(v)\left[1 - F(v)\right]^{n-1} \end{aligned}$$

$U_n = \max\{X_1, \cdots, X_n\}, X_i \sim unif(0,1)$

$$\begin{aligned} U_n &\sim Beta(n, 1) \\ f_n(u) &= nu^{n-1}, u \in [0,1] \\ F_n(u) &= u^n \\ E(U_n) &= \frac{n}{n+1} =: \mu_n \\ Var(U_n) &= \frac{n}{(n+1)^2(n+2)} =: \sigma_n^2 \end{aligned}$$

Note: convert any $unif(\theta - 1, \theta + 1)$ to $unif(0,1)$ and apply known knowledge

$V_n = \min\{X_1, \cdots, X_n\}, X_i \sim unif(0,1)$

$$\begin{aligned} V_n &\sim Beta(1, n) \\ f_n(v) &= n(1-v)^{n-1}, u \in [0,1] \\ F_n(v) &= 1 - (1-v)^n \\ E(V_n) &= \frac{1}{n+1} =: \mu_n \\ Var(V_n) &= \frac{n}{(n+1)^2(n+2)} =: \sigma_n^2 \end{aligned}$$

Limiting value for maximum

Note: this is not a question on central limit theorem

$$Z_n = \frac{U_n - \mu_n}{\sigma_n} = aU_n + b, a = \frac{1}{\sigma_n}, b = -\frac{\mu_n}{\sigma_n}$$

$$F_{Z_n}(z) = F_n(z/a - b/a) = \begin{cases} 0, & \mu_n + z\sigma_n < 0 \\ (\mu_n + z\sigma_n)^n, & 0 \le \mu_n + z\sigma_n \le 1 \\ 1, & \mu_n + z\sigma_n > 1 \end{cases}$$

$$\lim_{n\to\infty} F_{Z_n}(z) \to F_Z(z) = \begin{cases} e^{z-1}, & z \le 1 \\ 1, & z > 1 \end{cases}$$

$$\begin{aligned} \mu_n + z\sigma_n &= \frac{n}{n+1} + \frac{z}{n}\frac{n}{n+1}\sqrt{\frac{n}{n+2}} \\ &= (1 - \frac{1}{n}\frac{n}{n+1})(1 + \frac{z}{n}\sqrt{\frac{n}{n+2}}) \end{aligned}$$

$$\lim_{n\to\infty} (\mu_n + z\sigma_n)^n = e^{-1}\cdot e^z, z \le 1$$

MLE for maximum

consider i.i.d $X_1, X_2, \cdots, X_n \sim unif(0,\theta)$

$$\ell(\theta) = \begin{cases} -n\log(\theta), & 0 \le X_i \le \theta \ \forall i \\ -\infty, & \text{otherwise} \end{cases}$$

$$\Leftrightarrow \ell(\theta) = \begin{cases} -n\log(\theta), & \max\{X_1, \cdots, X_n\} \le \theta \\ -\infty, & \text{otherwise} \end{cases}$$

Since $\ell(\theta)$ is strictly decreasing function of θ ($\ell'(\theta) < 0$) for $\theta \ge \max\{X_1, \cdots, X_n\} (> 0)$, $\max \ell(\theta)$ at $\theta_{\min} = \max\{X_1, \cdots, X_n\}$
Basically, the smallest θ possible

Expected Values

Definition 4

$$E(X) = \begin{cases} \sum_i x_i p(x_i) \\ \int_{-\infty}^{\infty} x f(x)dx \end{cases}$$

Theorem 2: $Y = g(X)$

$$E(Y) = \begin{cases} \sum_i g(x_i)p(x_i) \\ \int_{-\infty}^{\infty} g(x)f(x)dx \end{cases}$$

Theorem 3: $Y = g(\mathbf{X}) = g(X_1, \cdots, X_n)$

$$E(Y) = \begin{cases} \sum_{x_1,\cdots,x_n} g(x_i)p(x_i) \\ \int\cdots\int g(x_i)f(x_i)dx_1\cdots dx_n \end{cases}$$

Corollary 1: X, Y are independent and g, h are fixed fn

$$E\left[g(X)h(y)\right] = E[g(X)] \cdot E[h(Y)]$$

Theorem 4: $Y = a + \sum_{i=1}^n b_i X_i$

$$E(Y) = a + \sum_{i=1}^n b_i E(X_i)$$

Even Odd function

For odd functions ($f_1(x) = xe^{-x^2/2}$), integral over a symmetric interval about 0 is zero.

$$\int_{-\infty}^{\infty} g_{odd}(x)dx = 0$$

For even functions ($f_0(x) = e^{-x^2/2}$, $f_2(x) = x^2e^{-x^2/2}$)

$$\int_{-\infty}^{\infty} g_{even}(x)dx = 2\cdot\int_0^{\infty} g_{even}(x)dx$$

Variance and Standard Deviation

Definition 5: $\mu = E(X)$

$$Var(X) = E[(X-\mu)^2] = \begin{cases} \sum_i (x_i - \mu)^2 p(x_i) \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \end{cases}$$

Theorem 5: if $Var(X)$ exist and $Y = a + bX$

$$Var(Y) = b^2 Var(X)$$

Theorem 6: if $Var(X)$ exist then

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

Theorem 7: Chebyshev's Inequality: for any $t > 0$

$$P(|X-\mu| > t) \le \frac{\sigma^2}{t^2}$$

if σ^2 is very small, there is a high probability that X will not deviate much from μ
Corollary 2:

$$Var(X) = 0 \Rightarrow P(X = \mu) = 1$$

Corollary 3: if X_i are independent

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i)$$

Moment-Generating Function

Definition 6 The moment generating function (mgf) of a RV X is

$$M(t) = E[e^{tX}] = \begin{cases} \sum_i e^{tx_i}p(x_i), & \text{[discrete]} \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & \text{[continuous]} \end{cases}$$

Limit Theorems

The Law of Large Numbers

Theorem 8 Let $X_1, \cdots, X_n$ be i.i.d RV with $E(X_i) = \mu, Var(X_i) = \sigma^2$. $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$
For any $\epsilon > 0$

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

From Chebyshev's inequality with $E(\bar{X}_n) = \mu, Var(\bar{X}_n) = \sigma^2/n$
Converge in probability to $\alpha \Leftrightarrow P(|Z_n - \alpha| > \epsilon) \to 0, n \to \infty$

Proving consistency with WLLN

Claim: σ^2 is consistently estimated by $(1/n)\sum_{i=1}^n X_i^2 - \bar{X}^2$

- From WWLN, $(1/n)\sum_{i=1}^n X_i^2 - \bar{X}^2$
- $\bar{X}^2 \to_p [E(X)]^2$
  ($Z_n \to_p \alpha \Rightarrow g(Z_n) \to_p g(\alpha)$ for any continuous g)
- $(1/n)\sum_{i=1}^n X_i^2 - \bar{X}^2 \to_p E(X^2) - [E(X)]^2 = Var(X)$

Convergence in Distribution

Definition 7 Let $X, X_1, X_2, \cdots$ be sequence of RV with cdf $F, F_1, F_2, \cdots$. X_n converges in distribution to X if

$$F_n(x) \to F(x), n \to \infty$$

for every cdf at every point at which F is continuous

Central Limit Theorem

Consider $X_1, X_2, \cdots$ sequence of i.i.d. with mean μ and variance σ^2. Let $S_n = \sum_{i=1}^n X_i$, then for $-\infty < x < \infty$

$$P(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x) \to \Phi(x), n \to \infty$$

CLT is concerned with how S_n/n fluctuates around μ

Sampling Distribution

χ^2 distribution

Note: χ^2 test is always right tailed

$$P\left(\chi_n^2(1-\alpha/2) \le x \le \chi_n^2(\alpha/2)\right)$$

Definition 8: if $Z \sim N(0,1)$, then

$$U = Z^2 \sim \chi_1^2,\ df = 1$$
$$f(u) = \frac{1}{\sqrt{2\pi}} u^{-1/2} e^{-u/2},\ u \ge 0$$
$$F(u) = \frac{1}{\sqrt{\pi}} \gamma(\frac{1}{2}, \frac{u}{2})$$
$$\chi_1^2 \sim \Gamma(\alpha = \frac{1}{2}, \lambda = \frac{1}{2})$$

Note: $\Gamma(1/2) = \sqrt{\pi}$

multiple $\chi_1^2 = \chi_n^2$

Definition 9: if $U_1, U_2, \cdots, U_n$ are independent χ_1^2, then

$$V = U_1 + U_2 + \cdots + U_n \sim \chi_n^2$$
$$f(v) = \frac{v^{n/2-1} e^{-v/2}}{2^{n/2}\Gamma(n/2)},\ v \ge 0$$
$$F(v) = \frac{1}{\Gamma(n/2)} \gamma(\frac{n}{2}, \frac{v}{2})$$
$$\chi_n^2 \sim \Gamma(\alpha = \frac{n}{2}, \lambda = \frac{1}{2})$$

Note: $E(V) = n, Var(V) = 2n$
if $U \sim \chi_m^2, V \sim \chi_n^2 \Rightarrow U + V \sim \chi_{m+n}^2$

t distribution

Definition 10: if $Z \sim N(0,1), U \sim \chi_n^2$ and Z, U independent

$$T = \frac{Z}{\sqrt{U/n}} \sim t_n,\ df = n$$

Proposition 3:

$$f(t) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2},\ -\infty < t < \infty$$

$E(T) = 0$ for df > 1, else undefined
$Var(T) = (df)/(df-2)$ for df > 2, ∞ if $1 <$ df ≤ 2, else undefined
Note:
t is symmetric about $0 \because f(t) = f(-t)$
t_1 is Cauchy distribution
$t_n \to N(0,1)$ as $n \to \infty$ (tail become lighter)

F distribution

Definition 11: if $U \sim \chi_m^2, V \sim \chi_n^2$, U, V independent

$$W = \frac{U/m}{V/n} \sim F_{m,n},\ df:\ m, n$$

Proposition 4: for $w \ge 0$

$$f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \frac{m}{n}^{m/2} w^{m/2-1} \left(1 + \frac{m}{n} w\right)^{-(m+n)/2}$$

$E(W) = n/(n-2), n > 2$
$Var(W) = (2n^2(m+n-2))/(m(n-2)^2(n-4))$
Note:
no $E(W)$ for $n \le 2$
$t_n^2 \sim F_{1,n}$

Double exponential (μ, λ)

$x \in \mathbb{R}$

$$f(x) = \frac{1}{2\lambda} exp(-\frac{|x-\mu|}{\lambda})$$
$$F(x) = \begin{cases} \frac{1}{2} exp(\frac{x-\mu}{\lambda}), & x \le \mu \\ 1 - \frac{1}{2} exp(-\frac{x-\mu}{\lambda}), & x \ge \mu \end{cases}$$

$E(X) = \mu, Var(X) = 2\lambda^2$

Beta (α, β)

$x \in [0,1]$

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$
$$F(x) = I_x(\alpha, \beta) = \frac{B(x;\alpha,\beta)}{B(\alpha,\beta)}$$
$$B(x;\alpha,\beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$$

$E(X) = \alpha/(\alpha+\beta)$
$Var(X) = (\alpha\beta)/((\alpha+\beta)^2(\alpha+\beta+1))$

Angular density (α)

Consider the angle θ at which electrons are emited in muon decay with $x \in [-1,1], \alpha \in [-1,1]$, $x = cos(\theta)$

$$f(x|\alpha) = \frac{1+\alpha x}{2}$$

$E(X) = \alpha/3$
$Var(X) = \frac{1}{3} - \frac{\alpha}{3}^3 = \frac{3-\alpha^2}{9}$

unknown dist

$x \in [0,1]$

$$f(x) = \theta x^{\theta-1}$$
$$F(x) = x^{\theta}$$

$E(x) = \frac{\theta}{\theta+1}$
$E(X^2) = \frac{\theta}{\theta+2}$
$Var(x) = -\frac{\theta}{(\theta+2)(\theta+1)}$

Sample Mean: $\bar{X}$, Sample Variance: S^2

Let $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ independently

$$\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$$
$$S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

Theorem 10:
$\bar{X}$ and $(X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X})$ are independent
Corollary 4:
$\bar{X}$ and S^2 are independent
Theorem 11:

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

Corollary 5:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Comparing variance estimates

Comparing

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$$
$$\tilde{\sigma}^2 = \rho\sum_{i=1}^n (X_i - \bar{X})^2$$

Now, since $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right] = n-1 \Rightarrow E(S^2) = \sigma^2$$

$$\hat{\sigma}^2 = \frac{n-1}{n}S^2 \Rightarrow E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

$$E(\tilde{\sigma}^2) = \rho(n-1)S^2 \Rightarrow \rho(n-1)\sigma^2$$

ρ that $\min MSE$ is $1/(n+1)$

Estimation of Parameters and Fitting of Distribution

Parmeter Estimation

For independent and identically distributed (i.i.d) RV

$$f(x_1, \cdots, x_n|\theta) = f(x_1|\theta)\cdots f(x_n|\theta)$$

An estimate of θ will be RV with sampling distribution.
Variability will be estimated through standard error, SE

The Method of Moments

Definition 12 population kth moment : $\mu_k = E(X^k)$

sample kth moment : $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$

Procedure to construct method of moments estimate

1. Express low-order moments in terms of the parameters
$$\mu_1 = E(X) = \mu, \mu_2 = E(X^2) = \mu^2 + \sigma^2$$
2. Invert to express the paraameters in terms of the moments
$$\Rightarrow \mu = \mu_1, \sigma^2 = \mu_2 - \mu_1^2$$
3. Insert sample moments to obtain estimate of the parameters
$$\Rightarrow \hat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{1}{n}X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$$

WLLN ensures that $\hat{\mu}_k \to_p \mu_k$
MoM is useful as the starting point for MLE estimation

δ method

for $\hat{\theta}_X = g(\bar{X})$

$$E(\hat{\theta}_X) \approx g[E(\bar{X})] + \frac{1}{2}g''[E(\bar{X})]Var(\bar{X})$$

$$Var(\hat{\theta}_X) \approx g'[E(\bar{X})]^2 Var(\bar{X})$$

Consistency

Definition 13: $\hat{\theta}_n$ is consistent in probability if $\hat{\theta}_n$ converges in probability to θ as $n \to \infty$. i.e. for any $\epsilon > 0$

$$P(|\hat{\theta}_n - \theta| > \epsilon) \to 0, n \to \infty$$

The Method of Maximum Likelihood

Definition 14: $f(\mathbf{X}|\theta) = f(x_1, \cdots, x_n|\theta)$
mle of θ is the value that $\max_\theta lik(\theta) = f(\mathbf{X}|\theta) \Leftrightarrow \max_\theta \ell(\theta)$

$$lik(\theta) = \prod_{i=1}^n f(X_i|\theta) = f(X_1|\theta)\cdots f(X_n|\theta)$$

$$\ell(\theta) = \sum_{i=1}^n \log[f(X_i|\theta)]$$

Note:
1. use $\ell(\theta) \doteq$ to omit the constant terms
2. eaiser to compute MLE for individual X_i and take sum
3. Sampling distribution of MLE are typically substantially less dispersed than MOM estimates. Therefore, more precise.

MLEs of multinomial cell probabilities

$$f(\mathbf{X}|p_1, \cdots, p_m) = \frac{n!}{x_1!\cdots x_m!}p_1^{x_1}\cdots p_m^{x_m}$$

$$\ell(p_1, \cdots, p_m) = \log n! - \sum_{i=1}^m logX_i! + \sum_{i=1}^m X_i \log p_i$$

in terms of other parameters

$$\ell(\theta) = \log n! - \sum_{i=1}^m logX_i! + \sum_{i=1}^m X_i \log p_i(\theta)$$

solve (Substitution)

$$p_m := 1 - \sum_{i=1}^{m-1} p_i$$

$$\ell(p_1, \cdots, p_{m-1}) \doteq \sum_{i=1}^{m-1} X_i \log(p_i) + X_m \log\left(1 - \sum_{i=1}^{m-1} p_i\right)$$

$$\frac{\partial \ell}{\partial p_j} = \frac{X_j}{p_j} - \frac{X_m}{p_m} = 0, j \in [1, m-1]$$

$$\Rightarrow \frac{X_1}{\hat{p}_1} = \frac{X_2}{\hat{p}_2} = \cdots = \frac{X_m}{\hat{p}_m} = \lambda = n$$

solve (Lagrange multiplier)

$$\max_{p_1, \cdots, p_m} \log n! - \sum_{i=1}^m logX_i! + \sum_{i=1}^m X_i \log p_i$$

$$s.t. \sum_{i=1}^m p_i = 1$$

result: $\hat{p}_j = \frac{X_j}{n}, j \in [1, m]$

MLE with param depending on θ

Suppose iid $X_1, X_2, \cdots, X_n \sim unif(0, \theta)$

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases}$$

$$log(f(x|\theta)) = \begin{cases} -log(\theta), & 0 \le x \le \theta \\ -\infty, & \text{otherwise} \end{cases}$$

$$\ell(\theta) = \begin{cases} -nlog(\theta), & 0 \le X_i \le \theta \ \forall i \\ -\infty, & \text{otherwise} \end{cases}$$

Fisher information

Fisher information (in one observation)

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

Lemma 1 under appropriate smoothness condition

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

Large sample theory for MLEs

Note: this is approximation using LLN
Theorem 12: under appropriate smoothness condition on f

1. the mle $\hat{\theta}$ from an i.i.d. sample is consistent
2. probability distribution $\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \to N(0, 1)$ where θ_0 is the true value of θ

Comments

- $\hat{\theta} \sim N(\theta_0, \frac{1}{nI(\theta_0)})$ for large sample
- mle is asymptotically unbiased
- asymptotic $= \lim_{n\to\infty} \frac{1}{nI(\theta_0)} = 0$ (very close)

- For i.i.d. sample size n
  Fisher information: $nI(\theta)$
  asymptotic variance: $1/[nI(\theta_0)]$
- For general sample size n
  Fisher information: $E[\ell(\theta)^2]$ or $-E[\ell''(\theta)]$
  asymptotic variance: $1/E[\ell'(\theta)^2]$ or $-1/E[\ell''(\theta)]$

Confidence intervals from MLEs

Definition 15: $100(1-\alpha)\%$ confidence interval for θ contains θ with probability $1-\alpha$. e.g. $\alpha = 0.05$ and CI = 95%
Want: (exact method)

$$P\left\{f(\frac{\alpha}{2}) \leq \mu \leq f(1-\frac{\alpha}{2})\right\} = 1-\alpha$$

Result:

$$\mu \in \bar{X} \pm \frac{S}{\sqrt{n}} t_{n-1}(\frac{\alpha}{2})$$
$$\sigma^2 \in \left(\frac{n\hat{\sigma}^2}{\chi^2_{n-1}(\alpha/2)}, \frac{n\hat{\sigma}^2}{\chi^2_{n-1}(1-\alpha/2)}\right)$$

Want: (approximate method)

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

Result:

$$\theta \in \hat{\theta} \pm z(\alpha/2)/\sqrt{nI(\hat{\theta})}$$

For multinomial (non i.i.d)

$$\theta \in \hat{\theta} \pm z(\alpha/2)/\sqrt{-E[\ell''(\hat{\theta})]}$$
$$\hat{\theta} \pm z(\alpha/2)\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{2n}}$$

Efficiency

Definition 16

1. mean squared error of $\hat{\theta}$

$$MSE(\hat{\theta}) = E[(\hat{\theta}-\theta_0)^2] = Var(\hat{\theta}) + [E(\hat{\theta})-\theta_0]^2$$

2. efficiency of $\hat{\theta}$ relative to $\tilde{\theta}$
   (both unbiased or has the same biased)

$$\text{eff}(\hat{\theta}, \tilde{\theta}) = \frac{Var(\tilde{\theta})}{Var(\hat{\theta})}$$

Cramer-Rao lower bound

Theorem 13
$T := t(X_1, \cdots, X_n)$ be unbiased estimate of θ

$$Var(T) \geq \frac{1}{nI(\theta)}$$
$$Var(T) \geq \frac{1}{I(\theta)} \text{ (multinomial)}$$

comments

- provides the lower bound on the variance of any unbiased estimate
- unbiased estimate achieve lower bound is efficient
- mle are asymptotically efficient as asymptotic variance = lower bound

Sufficiency

Definition 17
$T(X_1, \cdots, X_n)$ is sufficient for θ if the conditional distribution of $X_1, \cdots, X_n$ given $T = t$ does not depend on θ for any value of t.
T is called a sufficient statistic
Note: sufficiency is unique upto monotone transformation (e.g. $\log(x), x$)

A factorization theorem

Theorem 14
Express joint probability into functions containing only $\mathbf{X}$

$$f(\mathbf{X}|\theta) = g[T(\mathbf{X}), \theta]h(\mathbf{X}) \Leftrightarrow T(\mathbf{X}) \text{ is sufficient stat}$$

1. Identify joint probability function
2. Group terms into $g(t(x), \theta)h(x)$
3. $t(x)$ is the sufficient statistic

Corollary 6
If T is sufficient for θ, then the mle is a function of T
Note: to max MLE, it is sufficient to max T in this case.
This identify is useful for ratio test as well

$$\frac{lik(\theta_0)}{lik(\theta_1)} = \frac{g(T,\theta_0)h(x)}{g(T,\theta_1)h(x)} = \frac{g(T,\theta_0)}{g(T,\theta_1)}$$

Exponential family of distributions

RV with same dimension of "sufficient statistics" as "parameter space" regarldess of sample size
One parameter members (e.g. Ber, Binomial, Poisson)

$$f(x|\theta) = \begin{cases} exp[c(\theta)T(x) + d(\theta) + S(x)], & x \in A \\ 0, & x \notin A \end{cases}$$

k-parameter member (e.g. Normal, Gamma)

$$f(x|\theta) = \begin{cases} exp\left[\sum_{i=1}^k c_i(\theta)T_i(x) + d(\theta) + S(x)\right], & x \in A \\ 0, & x \notin A \end{cases}$$

where set A does not depend on θ

Checking exponential family

Since

$$\alpha^\beta = e^{\beta\log(\alpha)}$$

We can convert any function into a exp base. Therefore, taking $\log(f(x))$ and check which family dist belongs to

The Rao-Blackwell theorem

Theorem 15
$\hat{\theta}$ is estimator of θ, T is sufficient for θ, $\tilde{\theta} = E(\hat{\theta}|T)$

$$E[(\tilde{\theta}-\theta)^2] \leq E[(\hat{\theta}-\theta)^2]$$

If an estimator is not a function of a sufficient statistic, it can be improved

Testing Hypotheses and Assessing Goodness of Fit

Statistical hypothesis testing is a formal means of distinguishing between probability distributions on the basis of RV generated from one of the distribution
key idea: likelihood ratio

$$\frac{P(x|H_0)}{P(x|H_1)}$$

The Neyman-Pearson Paradigm

Hypothesis testing as a decision problem

H_0	: null hypothesis
H_1	: alternative hypothesis
Type I error	: rejecting H_0 when it is true
α	: significance level, probability of Type I error (e.g. 0.05)
Type II error, β	: accepting H_0 when it is false
Power, $1-\beta$	: probability of rejecting H_0 when it is false
test statistic	: likelihood ratio
rejection region	: set of values of test statistic leads to rejection of H_0
acceptance region	: set of values of test statistic lead to acceptance of H_0
null distribution	: probability distribution of test statistic when H_0 is true
simple hypothesis	: H_i completely specifies the probability distribution
composite hypothesis	: hypothesis does not completely specify the probability distribution

Theorem 16

Given simple hypotheses H_0, H_1 and test that reject H_0 with likelihood ratio $< c$ has significance level α

Then any other test with sifnificance level $\leq \alpha$ has power $\leq$ that of the likelihood ratio test

Or: Among all tests with given P(type I error), likelihood ratio test minimizes P(type II error)

Specifying the significance level and the concept of p-value

1. Specifying the significance level α

   Find α s.t. $P(|T \geq t_0|H_0) = \alpha$

2. Reporting the p-value

   summarise evidence against H_0 with p-value

   p-value = smallest sig level to reject H_0

The null hypothesis H_0

Asymmetry in the Neyman-Pearson paradigm between the null and alternative hypotheses

- Conventional to choose simpler hypotheses as null
- Choose hypothesis with greater consequences when incorrectly rejected (e.g. new drug).

  Because probability of rejecting can be controlled by α
- In scientific investigation, null hypothesis is simple explanation that must be discredited to demonstrate presence of a physical phenomenon or effect

Uniformly most powerful tests

Given a composite H_1, a uniformly most powerful test is one that is most powerful for every simple alternative H_1
E.g. happen when test does not depend on μ_1
Note: in typical composite situations, there is no uniformly most powerful test
Answering: The test is most powerful for testing $\lambda = \lambda_0$ vs $\lambda = \lambda_1 > \lambda_0$ and is the same for every such alternative

The Duality of Confidence Intervals and Hypothesis Tests

Inversion: confidence set can be obtained by "inverting" a hypothesis test, and vice versa

Theorem 17

Suppose that for every value θ_0 in Θ there is a test at level α of the hypothesis $H_0 : \theta = \theta_0$ with acceptance region $A(\theta_0)$. Then the set

$$C(\mathbf{X}) = \{\theta : \mathbf{X} \in A(\theta)\}$$

is a $100(1-\alpha)\%$ confidence region for θ
In words: A $100(1-\alpha)\%$ confidence region for θ consists of those values of θ_0 for which $H_0 : \theta = \theta_0$ will not be rejected at level α

Theorem 18

Suppose that $C(\mathbf{X})$ is a $100(1-\alpha)\%$ confidence region for θ. Then an acceptance region for a level α test of the hypothesis $H_0 : \theta = \theta_0$ is

$$A(\theta_0) = \{\mathbf{X} : \theta_0 \in C(\mathbf{X})\}$$

In words: The hypothesis that $\theta = \theta_0$ is accepted if θ_0 lies in the confidence region.

Generalized Likelihood Ratio Tests

Given $\mathbf{X} = (X_1, \cdots, X_n)$ with $f(\mathbf{X}|\theta)$.
Let ω_0, ω_1 be subsets of all possible values of θ s.t. ω_1 is disjoint from ω_0 and $\Omega = \omega_0 \cup \omega_1$
For testing $H_0 : \theta \in \omega_0$ v.s. $H_1 : \theta \in \omega_1$

$$\Lambda = \frac{\max_{\theta \in \omega_0} lik(\theta)}{\max_{\theta \in \Omega} lik(\theta)}$$

Reject H_0 for a small Λ
$S = \{x : T(x) > / < c\}$, $P(S) = \alpha$

Theorem 19

Under smoothness conditions on the probability density or frequency functions involved

$$-2\log(\Lambda) \sim \chi^2_{df}, n \to \infty$$

df = dim Ω - dim ω_0
Reject H_0 for large $-2\log\Lambda > \chi^2_{df}(\alpha)$

degree of freedom: number of free parameters under Ω and ω_0 respectively.
e.g. $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$, df $= 1 - 0$
μ is specified under H_0 but needs to be esitmated under H_1

General steps for Ratio test

Refer: problem 50

1. Identify hypothesis as simple/composite

   [simple] subsitute into lik

   [composite] find MLE estimate

2. Set up likelihood ratio and find Λ

   $$\Lambda = \frac{f(\mathbf{X}|H_0)}{f(\mathbf{X}|H_1)}$$

3. Find extreme values (c) that min Λ and reject H_0

   [max] $P(g(T(\mathbf{X}), \theta) > c|H_0) = \alpha$

   [min] $P(g(T(\mathbf{X}), \theta) < c|H_0) = \alpha$

4. Often, find $T(\mathbf{X})$ is easier

   [one tail] $P(T(\mathbf{X}) > c) = \alpha$

   [two tail] $P(-c < T(\mathbf{X}) < c) = \alpha$

5. If exact Λ is hard to find, use $-2\log(\Lambda) \sim \chi^2_{df}$ by large sample approx

Likelihood Ratio Tests for the Multinomial Distribution

$H_0 : p = p(\theta), \theta \in \omega_0$
e.g. λ in Pois

$$-2\log(\Lambda) = 2\sum_{i=1}^{m} O_i \log\left(\frac{O_i}{E_i}\right)$$

X^2 and $-2\log(\Lambda)$ are asymptotically equivalent under H_0

Pearson's χ^2 statistics

Pearson's chi-square statistic (assess goodness of fit)

$$X^2 = \sum_{\text{all cells}} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{df}, n \to \infty$$

$O_i :=$ observed count
$E_i :=$ expected count
df := #cell - #independent parameters - 1

Require expected counts ≥ 5

Investigate when goodness-of-fit test failed

Look for cells that make large contributions to X^2 and note whether $O > E$ or $O < E$

Comparing Two Samples

In many experiments, the two samples maybe regarded as being independent of each other.

Only continuous measurements and parametric methods are discussed in this module

Comparing Two Independent Samples

Model:

- Observations from control group are independent RV with common distribution F
- Treatment group are independent RV with common distribution G

Objective: inference about the comparison of F, G (usually difference of means) based on normal distribution

Methods based on Normal distribution

Note:

- mle of $\mu_X - \mu_Y = \bar{X} - \bar{Y}$
- $\bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \sigma^2\left[\frac{1}{n} + \frac{1}{m}\right])$
- If σ^2 is known

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{1/n + 1/m}} \sim N(0,1)$$

- If σ^2 is unknown, it can be estimated with pooled sample variance

$$s_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{m+n-2}$$
$$\hat{\sigma} = s_{\bar{X}-\bar{Y}} = s_p\sqrt{1/n + 1/m}$$

Theorem 20
Supposed that Xs are independent of Ys with
iid $X_i \sim N(\mu_X, \sigma^2)$, $i \in [1, n]$ and
iid $Y_j \sim N(\mu_Y, \sigma^2)$, $j \in [1, m]$

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{s_{\bar{X}-\bar{Y}}} \sim t_{df}$$

$df = m + n - 2$

Corollary 7
Under assumptions of Theorem 20, a $100(1-\alpha)\%$ CI for $\mu_X - \mu_Y$ is

$$(\bar{X} - \bar{Y}) \pm t_{m+n-2}(\alpha/2) s_{\bar{X}-\bar{Y}}$$

Test for unequal variance

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} = \frac{\sum_{i=1}^{n}(X_i - \hat{\mu}_0)^2 + \sum_{j=1}^{m}(Y_j - \hat{\mu}_0)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{j=1}^{m}(Y_j - \bar{Y})^2} \sim |t|$$

Unequal variance

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n + S_Y^2/m}}$$
$$df = \frac{(S_X^2/n + S_Y^2/m)^2}{(S_X^2/n)^2/(n-1) + (S_Y^2/m)^2/(m-1)}$$

Specific Example questions

Capture/Recapture Method

Known $t :=$ number of animals captured, tagged, and released
$m :=$ number of animals captured in the second try
$r :=$ number of animals tagged (in second capture)
Interested to know the size of population (n)

$$L_n = \binom{t}{r}\binom{n-t}{m-r} / \binom{n}{m}$$

$L_n :=$ probability of r capture, assuming equal probability among $\binom{n}{m}$ groups
Solution: max integer s.t. $n < mt/r$

Discrete RV

Given sample space $\Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$
$X :=$ number of heads of first toss
$Y :=$ total number of heads

$p(x,y)$	$y=0$	$y=1$	$y=2$	$y=3$	$\to$	$p_X(x)$
$x=0$	1/8	2/8	1/8	0	$\to$	1/2
$x=1$	0	1/8	2/8	1/8	$\to$	1/2
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\to$	
$p_Y(y)$	1/8	3/8	3/8	1/8	$\to$	

cell shows the joint frequency function
summing across the rows and columns will get the marginal frequency functions.

Finding pivot, exact CI

Tutorial 6: consider $\bar{Y} \sim \Gamma(\alpha = n, \lambda = n\theta)$

$$2n\theta\bar{Y} \sim \Gamma(\alpha = n, \lambda = 1/2) \Leftrightarrow \chi^2_{2n}$$

$2n\theta\bar{Y}$ is a pivot

$$1 - \alpha = P\{\chi^2_{2n}(1-\alpha/2) \leq 2n\theta\bar{Y} \leq \chi^2_{2n}(\alpha/2)\}$$
$$= P\left\{\frac{\chi^2_{2n}(1-\alpha/2)}{2n\bar{Y}} \leq \theta \leq \frac{\chi^2_{2n}(\alpha/2)}{2n\bar{Y}}\right\}$$

Twins

Reference: Problem 8, 36, 39

Find distribution

Problem 8: In the population of twins, male (M) and females (F) are equal likely to occur and probability of identical twins are α. If twins are not identical, their genes are independent.

Let $B_1 :=$ identical twins, $B_2 :=$ non identical twins

$$
\begin{aligned}
P(MM) &= P(MM|B_1)P(B_1) + P(MM|B_2)P(B_2) \\
&= \frac{1}{2}\alpha + (\frac{1}{2}\cdot\frac{1}{2})(1-\alpha) \\
&= \frac{1+\alpha}{4} = P(FF) \\
P(MF) &= 1 - P(MM) - P(FF) = \frac{1-\alpha}{2} \\
p_1(\alpha) &= P(MM) = \frac{1+\alpha}{4} \\
p_2(\alpha) &= P(FF) = \frac{1+\alpha}{4} \\
p_3(\alpha) &= P(MF) = \frac{1-\alpha}{2}
\end{aligned}
$$

Find MLE

Problem 36: Supposed n twins are sampled. n_1 are MM, n_2 are FF, n_3 are MF. But unknown which tiwns are identical.

Find mle of α

$$
f(X_1, X_2, X_3|\alpha) = \binom{n}{X_1, X_2, X_3} p_1(\alpha)^{X_1} p_2(\alpha)^{X_2} p_3(\alpha)^{X_3}
$$

$$
\begin{aligned}
\ell(\alpha) &\doteq X_1 \log p_1(\alpha) + X_2 \log p_2(\alpha) + X_3 \log p_3(\alpha) \\
&\doteq X_1 \log(1-\alpha) + X_2 \log(1+\alpha) + X_3 \log(1-\alpha) \\
&= (X_1 + X_2)\log(1+\alpha) + X_3 \log(1-\alpha) \\
\ell'(\alpha) &= \frac{X_1 + X_2 - X_3 - n\alpha}{(1+\alpha)(1-\alpha)}
\end{aligned}
$$

Since $X_1 + X_2 + X_3 = n$, if $X_1 + X_2 - X_3 < 0 \Rightarrow \ell'(\alpha) < 0$

$$
\Rightarrow \hat{\alpha} = \begin{cases} 0, & X_1 + X_2 - X_3 < 0 \\ (X_1 + X_2 - X_3)/n, & \text{otherwise} \end{cases}
$$

Find asymptotic variance of MLE

$$
\begin{aligned}
\ell(\theta) &= X_1 \log(p_1) + X_2 \log(p_2) + X_3 \log(p_3) \\
\ell'(\theta) &= \frac{X_1 + X_2}{1-\alpha} - \frac{X_3}{1-\alpha} \\
\ell''(\theta) &= -\frac{X_1 + X_2}{(1+\alpha)^2} - \frac{X_3}{(1-\alpha)^2} \\
I(\theta) &= -E[\ell''(\theta)] = \frac{n}{1-\alpha}
\end{aligned}
$$

Hardy-Weinberg Law

Reference: Problem 9, 40
In general, questions like this

1. find the conditional probability
2. using Law of Total Probability to find the exact probability

Find distribution

Problem 9: Assumes genes can either be a, A. The possible genotypes are AA, Aa, aa.
When two organisms mate, each independently contribute one of genes with probability $p, 2q, r$ respectively

1st generation	probability	2nd generation
$B_1 = \{AA_1, AA_1\}$	$P(B_1) = p^2$	$P(AA_2\|B_1) = 1$
$B_2 = \{AA_1, Aa_1\}$	$P(B_2) = 2pq$	$P(AA_2\|B_2) = 0.5$ $P(Aa_2\|B_2) = 0.5$
$B_3 = \{AA_1, aa_1\}$	$P(B_3) = pr$	$P(Aa_2\|B_3) = 1$
$B_4 = \{Aa_1, AA_1\}$	same B_2	
$B_5 = \{Aa_1, Aa_a\}$	$P(B_5) = (2q)^2$	$P(AA_2\|B_5) = 0.25$ $P(Aa_2\|B_5) = 0.5$ $P(aa_2\|B_5) = 0.25$
$B_6 = \{Aa_1, aa_1\}$	$P(B_6) = 2qr$	$P(Aa_2\|B_6) = 0.5$ $P(aa_2\|B_6) = 0.5$
$B_7 = \{aa_1, AA_1\}$	same as B_3	
$B_8 = \{aa_1, Aa_1\}$	same as B_6	
$B_9 = \{aa_1, aa_1\}$	$P(B_9) = r^2$	$P(aa_2\|B_9) = 1$

with $\theta = q + r, 1 - \theta = p + q$

$$
\begin{aligned}
P(AA_2) &= \sum_{i=1}^{9} P(AA_2|B_i)P(B_i) = (p+q)^2 \\
&= (1-\theta)^2 \\
P(Aa_2) &= 2(p+q)(q+r) \\
&= 2\theta(1-\theta) \\
P(aa_2) &= (q+r)^2 \\
&= \theta^2
\end{aligned}
$$

with $p' = (1-\theta)^2, q' = \theta(1-\theta), r' = \theta^2$

$$
\begin{aligned}
P(AA_3) &= (p' + q')^2 = (1-\theta)^2 \\
P(Aa_3) &= 2(p' + q')(q' + r') = 2\theta(1-\theta) \\
P(aa_3) &= (q' + r')^2 = \theta^2
\end{aligned}
$$

Find MLE

If gene frequencies are in equilibrium, the genotypes AA, Aa, aa occur with probabilities $(1-\theta)^2, 2\theta(1-\theta)$ and θ^2 respectively.

$$
f(X_1, X_2, X_3|\theta) = \binom{n}{X_1, X_2, X_3} [(1-\theta)^2]^{X_1} [2\theta(1-\theta)]^{X_2} [\theta^2]^{X_3}
$$

$$
\begin{aligned}
\ell(\theta) &\doteq X_1 \log(1-\theta)^2 + X_2 \log 2\theta(1-\theta) + X_3 \log(\theta^2) \\
&\doteq (2X_1 + X_2)\log(1-\theta) + (X_2 + 2X_3)\log(\theta) \\
\hat{\theta} &= \frac{X_2 + 2X_3}{2n}
\end{aligned}
$$

Find asymptotic variance of MLE

$$
\begin{aligned}
\ell(\theta) &= X_1 \log(p_1) + X_2 \log(p_2) + X_3 \log(p_3) \\
\ell'(\theta) &= -\frac{2X_1 + X_2}{1-\theta} + \frac{2X_3 + X_2}{\theta} \\
\ell''(\theta) &= -\frac{2X_1 + X_2}{(1-\theta)^2} - \frac{2X_3 + X_2}{\theta^2} \\
I(\theta) &= -E(\ell''(\theta)) = \frac{2n}{(1-\theta)\theta}
\end{aligned}
$$

Distribution	Parameters (θ)	MOM	MLE	Fisher information $I(\theta)$	MLE asymptotic variance	Sufficient statistics $T(\mathbf{X})$	question ref
Discrete Distribution (i.i.d.)							
Bernoulli	p	$\hat{p} = \bar{X}$	$\hat{p} = \bar{X}$	$1/pq$	pq/n	$\sum_{i=1}^{n} X_i$	suff: ex26, 27 fam: ex29
Poisson	λ	$\hat{\lambda} = \bar{X}$	$\hat{\lambda} = \bar{X}$	$1/\lambda$	λ/n	$\sum_{i=1}^{n} X_i$	MOM: ex12 MLE: ex16 eff: ex25 htest: q49, ex31
Geometric	p	$\hat{p} = 1/\bar{X}$	$\hat{p} = 1/\bar{X}$	$1/\left[(p^2(1-p))\right]$	$p^2(1-p)/n$	$\sum_{i=1}^{n}(k_i - 1)$	MOM: q29 MLE: q33 MLE var: q37
Multinomial (Discrete, not independent)							
Binomial	p	$\hat{p} = X/n$	$\hat{p} = X/n$	$\frac{n}{p(1-p)}$	$\frac{p(1-p)}{n}$	X	MLE: q32 eff: q43 htest: q48, q56 q58
Negative Binomial (note: pmf diff from wiki)	p	$p = 1 - \frac{E(X)}{Var(X)}$	$\hat{p} = \frac{r}{r+k}$	$\frac{n}{p^2(1-p)}$	$\frac{p^2(1-p)}{n}$	X	
twins	α	-	$\max\{0, \frac{X_1+X_2-X_3}{n}\}$	$n/(1-\alpha^2)$	$(1-\alpha^2)/n$	X_1, X_2, X_3	MLE: q36 var: p39
H-W equilibrium	θ	-	$\hat{\theta} = (X_2 + 2X_3)/2n$	$-2n/[\theta(1-\theta)]$	$\theta(1-\theta)/2n$	X_1, X_2, X_3	MLE: ex20 var: ex23 htest: ex37, q59
cell probabilities	p_i	-	$\hat{p}_i = X_i/n$ -	-	-	-	

Distribution	Parameters (θ)	MOM	MLE	Fisher information $I(\theta)$	MLE asymptotic variance	Sufficient statistics $T(\mathbf{X})$	question ref
Continuous Distribution (i.i.d)							
Uniform $[0, \theta]$	θ	$\hat{\theta} = 2\bar{X}$	$\hat{\theta} = \max\{X_1, \cdots, X_n\}$	-	-	-	MOM: q31 MLE: q35 eff: q41 htest: q57
Uniform $[\theta-1, \theta+1]$	θ	$\hat{\theta} = \bar{X}$	$\hat{\theta} = X_i$, any i	-	-	-	MLE: q35
$f(x\|\theta) = \theta x^{\theta-1}$	θ	$\hat{\theta} = \bar{X}/(1-\bar{X})$	$\hat{\theta} = -n/(\sum_{i=1}^n log(X_i))$	n/θ^2	θ^2/n^2	$\prod_{i=1}^n X_i$	MOM: tut4
Exponential	λ	$\hat{\lambda} = 1/\bar{X}$	$\hat{\lambda} = 1/\bar{X}$	n/λ^2	λ^2/n^2	$\sum_{i=1}^n X_i$	MOM: tut4 MLE: tut5 E, Var: tut5 suff: q45 htest: q55
Double exponential [scale]	σ	$\hat{\sigma} = \sqrt{\hat{\mu}_2/2}$	$\hat{\sigma} = \frac{1}{n}\sum_{i=1}^n \|X_i\|$	$-1/\sigma^2$	σ^2/n	$\sum_{i=1}^n \|x_i\|$	MOM: q30 MLE: q34 var: p38 suff:q44
Gamma	α, λ	$\hat{\lambda} = \bar{X}/\hat{\sigma}^2$ $\hat{\alpha} = \bar{X}^2/\hat{\sigma}^2$ $\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2$	$\hat{\lambda} = \hat{\alpha}\bar{X}$ $\hat{\alpha}$: $n\log(\hat{\alpha}) - n\log(\bar{X})$ $+\sum_{i=1}^n \log(X_i) - \frac{n\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = 0$	$n\alpha\lambda^2$	$1/(n\alpha\lambda^2)$	$\prod_{i=1}^n X_i$ $\sum_{i=1}^n X_i$	MOM: ex14 MLE: ex18 suff: q46 fam:47
Normal	μ, σ^2	$\hat{\mu} = \bar{X}$ $\hat{\sigma} = \hat{\mu}_2 - \bar{X}^2$	$\hat{\mu} = \bar{X}$ $\hat{\sigma}^2 = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}$	$n/(2\sigma^2)$	$(2\sigma^2)/n$	$\prod_{i=1}^n X_i$ $\sum_{i=1}^n X_i$	MOM: ex13 MLE: ex17 eff: q42 suff: ex28 htest: ex30, q52 ex35
Angular [muon decay]	α	$\hat{\alpha} = 3\bar{X}$	$\hat{\alpha}$: $\sum_{i=1}^n X_i/(1+\hat{\alpha}X_i) = 0$	$(n\alpha)/(3-\alpha^2)$	$(3-\alpha^2)/(n\alpha)$	-	MOM ex15 MLE: ex19 E,Var :q28 eff: ex24
Beta	α, β	$\hat{\alpha} = \bar{X}\left[\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right]$ $\hat{\beta} = (1-\bar{X})\left[\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right]$	-	-	-	$\prod_{i=1}^n X_i$ $\prod_{i=1}^n (1-X_i)$	