UC Berkeley, Department of Statistics

MA Exam, (Statistics 201B)

January 21, 2023

1. Consider a clinical trial in which one group of patients receives an experimental cancer treatment and another group receives a standard treatment. The experimental group contains n patients, of which X are still alive five years later. The standard treatment group contains m patients, of which Y are still alive five years later. Let p_e denote the probability of surviving five years or more under the experimental treatment, and let p_s denote the same probability for the standard treatment, and model $X \sim \text{Binomial}(n; p_e)$ and $Y \sim \text{Binomial}(m; p_s)$. We assume that X and Y are independent.

   (a) Derive the MLEs for p_e and p_s, and for $\psi = p_e - p_s$.

   (b) Calculate the bias and variance for the MLE $\hat{\psi}$.

   (c) Find the Fisher information for ψ. Does it correspond to the variance of $\hat{\psi}$?

   (d) Consider testing $H_0 : p_e = p_s$ versus $H_1 : H_0$ is not true. Determine what test statistic you would use. What is the (asymptotic) sampling distribution of the test statistic? Set up the decision rule.

2. Suppose that $X_1, \ldots, X_n$ are i.i.d. Exponential(θ)

   (a) Find the MLE for $\log \theta$ and its variance.

   (b) Using a Gamma(α, β) prior, what is the Bayes estimator for θ under squared error loss?

   (c) In words (you do not need to carry out the mathematics), how would you decide between modeling these data as Exponential(θ) and Gamma(δ, σ)[1]?

[1] That is a Gamma model for the data distribution, not the prior used in part (b).

Distributions

Name/Range	Notation	pmf/pdf	Mean	Variance
Normal $(-\infty,\infty)$	$N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{(x-\mu)^2/2\sigma^2}$	$E\mu$	$\mathrm{Var}(X)=\sigma^2$ $\mathrm{Var}([X-\mu]^2)=2\sigma^4$
Exponential $(0,\infty)$	$Exp(\theta)$	$\theta e^{-\theta x}$	$1/\theta$	$1/\theta^2$
Uniform (a,b)	$U(a,b)$	$I(a\le x\le b)$	$(a+b)/2$	$(b-a)^2/12$
Gamma $(0,\infty)$	$G(\alpha,\beta)$	$\beta^\alpha x^{\alpha-1}e^{-\beta x}/\Gamma(\alpha)$	α/β	α/β^2
Beta $(0,1)$	$Beta(\alpha,\beta)$	$x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta)$	$\alpha/(\alpha+\beta)$	$\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$
Dirichlet	$D(\alpha)$	$B(\alpha)^{-1}\prod x_k^{\alpha_k-1}$	α_k/α_0	$(\alpha_k/\alpha_0)(1-\alpha_k/\alpha_0)/(\alpha_0+1)$
Binomial n tries, # success	$B(n,p)$	$\binom{n}{k}p^x(1-p)^{n-x}$	np	$np(1-p)$
Multinomial $x_k=$ no in cat k	$M(n,p)$	$\frac{n!}{x_1!\ldots x_k!}p_1^{x_1}\ldots p_k^{x_k}$	np_j	$np_j(1-p_j)$
Poisson +ive integers	$P(\lambda)$	$\lambda^k e^{-\lambda}/k!$	λ	λ
Negative binomial tries until r success	$NB(p,r)$	$\binom{k+r-1}{r-1}(1-p)^k p^r$	$r(1-p)/p$	$r(1-p)/p^2$

Results $X\sim(\mu,\sigma^2)$ means $EX=\mu$, $\mathrm{var}(X)=\sigma^2$ without assuming normality.

- Properties of expectation and variance

  $$E(aX+b)=a\mu+b \qquad \mathrm{var}(aX+b)=a^2\sigma^2 \qquad \mathrm{cov}(aX+Z,cY+d)=ac\,\mathrm{cov}(X,Y)+c\,\mathrm{cov}(Z,Y)$$

- Bonferroni's inequality: $P(A \text{ or } B)<P(A)+P(B)$.
- Probability limits if $X_i\sim(\mu,\sigma^2)$

  $$P(|\bar{X}_n-\mu|>\epsilon)\to 0 \qquad \sqrt{n}(\bar{X}_n-\mu)/\sigma\xrightarrow{d}N(0,1)$$

- Delta method, for a sequence Z_n (usually estimators)

  $$\sqrt{n}(Z_n-\mu)/\sigma\xrightarrow{d}N(0,1)\Rightarrow\sqrt{n}(g(Z_n)-g(\mu))/(g'(\mu)\sigma)\xrightarrow{d}N(0,1)$$

- M-estimator, if $X_1,\ldots,X_n$ i.i.d. and $\theta=\mathrm{argmax}s\sum M(\theta;X_i)$ then

  $$\sqrt{n}(\hat{\theta}-\theta)\approx\frac{\frac{1}{\sqrt{n}}\sum dM(\theta;X_i)/d\theta}{\frac{1}{n}\sum d^2M(\theta;X_i)/d\theta^2}\xrightarrow{d}N\left(0,\mathrm{var}(dM/d\theta)/\left[Ed^2M/d\theta^2\right]^2\right)$$

  When M is likelihood, $\mathrm{var}(dM/d\theta)=-Ed^2M/d\theta^2=I(\theta)$.
- MSE = bias^2 + variance: $E(X-a)^2=E(X-EX)^2+(EX-a)^2$
- Linear Regression estimates $\hat{\beta}_1=\sum(X_i-\bar{X})(Y_i-\bar{Y})/\sum(X_i-\bar{X})^2$, $\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}$.

Definitions

If $X_i \sim f(x; \theta^*)$

log likelihood $l_n(\theta; \mathcal{X}) = \sum \log f(X_i; \theta)$

score $s_n(\theta; \mathcal{X}) = dl_n(\theta; \mathcal{X})/d\theta$, note $E_{\theta^*} s_n(\theta^*; \mathcal{X}) = 0$.

information $I(\theta) = E_{\theta^*} \left[\frac{d}{d\theta} \log f(X, \theta^*)\right]^2 = -E_{\theta^*} \frac{d^2}{d\theta^2} \log f(X, \theta^*)$

MLE $\hat{\theta}_n = \text{argmax}_\theta \; l_n(\theta; \mathcal{X}) = \{\theta : s_n(\theta; \mathcal{X}) = 0\}$, $\sqrt{nI(\theta)}(\hat{\theta}_n - \theta^*) \xrightarrow{d} N(0, 1)$

MoM If $EX^j = \mu_j(\theta)$ then $\hat{\theta}_n$ solves $\mu_j(\hat{\theta}_n) = \frac{1}{n} \sum X_i^j$.

Empirical Distribution $\hat{F}_n(x; \mathcal{X}) = \frac{1}{n} \sum I(X_i < x)$, "draw an observed X_i at random"

LR Test Of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ rejects for $l(\theta_1; \mathcal{X})/l(\theta_0; \mathcal{X}) > C$.

Generalized LR Test for $\theta = \theta_0$: $2(\max l_n(\theta; \mathcal{X}) - l_n(\theta_0; \mathcal{X})) \sim \chi_1^2$

Wald test $\sqrt{nI(\theta)}|\hat{\theta} - \theta_0| < z_{1-\alpha/2}$

Bonferroni Correction For K tests, reject test with $p < \alpha/K$

Posterior $\mathcal{X} \sim f(\mathcal{X}|\theta)$, $\theta \sim f(\theta)$, posterior is $f(\theta|\mathcal{X}) \sim C(\mathcal{X}) f(\mathcal{X}|\theta) f(\theta)$.

Expected *a posteriori* $\hat{\theta} = \int \theta f(\theta|\mathcal{X}) d\theta$.

Bayes Factor between model $f_0(\mathcal{X}|\theta_0)$ with prior $f(\theta_0)$ and $f_1(\mathcal{X}|\theta_1)$ with prior $f(\theta_1)$ is $\int f(\mathcal{X}|\theta_0) f(\theta_0) d\theta_0 / \int f(\mathcal{X}|\theta_1) f(\theta_1) d\theta_1$.

Frequentist Risk for estimator $\hat{\theta}(\mathcal{X})$, $R(\theta, \hat{\theta})) = E_{\mathcal{X}} L(\theta, \hat{\theta}(\mathcal{X}))$

Posterior Risk $r(\hat{\theta}|\mathcal{X}) = \int L(\theta, \hat{\theta}(\mathcal{X})) f(\theta|\mathcal{X}) d\theta$

Bayes Estimator $\hat{\theta}(\mathcal{X}) = \text{argmin}_t r(t|\mathcal{X})$.

Bayes Risk $r(f, \hat{\theta}) = \int R(\theta, \hat{\theta}) f(\theta) d\theta$

Kernel (local) Estimate $\hat{\theta}(x) = \text{argmin}_t \sum K(X_i, x) L(t; X_i, Y_i)$ for squared error loss:

$$\hat{f}(x) = \frac{\sum_{i=1}^n K(|X_i - x|/h) Y_i}{\sum K(|X_i - x|/h)}$$

Basis Expansion $\hat{g}(x) = \sum_{j=1}^J b_j \phi_j(x)$