Statistics A exam solutions (2003-2004)

David Rossell i Ribera

18th January 2005

I may have committed a number of errors in writing these solutions, but they should be fine for the most part.Use at your own risk!

Statistics May 2003

Problem 1

a) $g(p)$ is U-estimable iff $\sum_{x=0}^{n} h(x) \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} = g(p)$. The left hand side is a sum of polynomial of degree $\leq n$, and hence the right hand side has to be a polynomial of degree $\leq n$.

b) From section a) we know that $h(x)$ has to be such that the sumation equals $p(1-p)^2 = p - 2p^2 + p^3$. Hence it suffices to choose $h(x)$ such that the coefficient affecting p in the left hand side is 1, for p^2 it's -2 and for p^3 it's 1. To do this we need to get through some messy algebra to expand the left hand side expression.

prova prova

Hint: $(1-p)^{n-x} = (1+(-p))^{n-x} = \sum_{k=0}^{n-x} \begin{pmatrix} n-x \\ k \end{pmatrix} 1^k (-p)^{n-x-k}$

so $\sum_{x=0}^{n} h(x) \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \sum_{k=0}^{n-x} h(x) \begin{pmatrix} n \\ x \end{pmatrix} \begin{pmatrix} n-x \\ k \end{pmatrix} (-1)^{n-k} p^{n-k}$

c) $g(x) = sin^{-1}(\sqrt{x})$ gives $g'(x) = \frac{1}{2\sqrt{x}cos(sin^{-1}(x))}$, and thus $(g'(x))^2 = \frac{1}{4xcos^2(sin^{-1}(x))} = \frac{1}{4x(1-sin^2(sin^{-1}(x)))} = \frac{1}{4x\left(1-(\sqrt{x})^2\right)} = \frac{1}{4x(1-x)}$

The delta method gives that $\sqrt{n}\left(sin^{-1}\left(\sqrt{x/n}\right) - sin^{-1}\left(\sqrt{p}\right)\right) \to^D N\left(0, \frac{1}{4}\right)$.

Problem 2

a) Note that $\int f_\theta(x)d\mu(x) = 1 \Rightarrow \int h(x)e^{\theta'T}d\mu(x) = e^{A(\theta)}$, which is convex. Hence $e^{A(\alpha\theta_1+(1-\alpha)\theta_2)} \leq \alpha e^{A(\theta_1)} + (1-\alpha)e^{A(\theta_2)} < \infty$.

b) First, let's derivate w.r.t. θ_i. $\int h(x)e^{\theta'T}d\mu(x) = e^{A(\theta)} \Rightarrow \int h(x)e^{\theta'T}T_i(x)d\mu(x) = e^{A(\theta)}\frac{d}{d\theta_i}A(\theta) \Rightarrow \frac{d}{d\theta_i}A(\theta) = \int h(x)e^{\theta'T-A(\theta)}T_i(x)d\mu(x) = E(T_i(X))$.

Now let's derivate again w.r.t. θ_j. $\frac{d^2}{d\theta_i d\theta_j}A(\theta) = \frac{d}{d\theta_j}\int h(x)e^{\theta'T-A(\theta)}T_i(x)d\mu(x) = \int h(x)e^{\theta'T-A(\theta)}(T_j(x) - \frac{d}{d\theta_j}A(\theta))T_i(x)d\mu(x) = E(T_i(X)T_j(X)) - E(T_i(X))E(T_j(X)) = COV(T_i(X), T_j(X))$.

c) The likelihood is $f_p(x) = \frac{n!}{\prod x_i!}exp\left\{\sum_{i=0}^r x_i log p_i\right\} = \frac{n!}{\prod x_i!}exp\left\{nlogp_0 + \sum_{i=0}^r x_i log\frac{p_i}{p_0}\right\}$. Now let $\eta_i = log\frac{p_i}{p_0} \Rightarrow p_0 = (1 + \sum_{i=1}^r e^{\eta_i})$, so we get

$f_\eta(x) = \frac{n!}{\prod x_i!}exp\left\{\sum x_i\eta_i - nlog(1+\sum e^{\eta_i})\right\}$. Hence $A(\eta) = nlog(1+\sum e^{\eta_i}) \Rightarrow \frac{d}{d\eta_i}A(\eta) = \frac{ne^{\eta_i}}{1+\sum e^{\eta_i}} = np_i$.

And so on...

Problem 3

a) Easy

b) $E(\hat{\sigma}^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{1}{n}E((n-1)S^2) = \frac{1}{n}\sigma^2(n-1)$, so $S^2 = \frac{n}{n-1}\hat{\sigma}^2$ is unbiased and a function of the complete & sufficient statistic $\Rightarrow$ UMVUE by Lehmann-Scheffe

c) $f(\mu, \sigma^2) \propto (2\tau)^n e^{-\tau\sum(x_i-\mu)^2}\frac{\alpha^g}{\Gamma(g)}\tau^{g-1}e^{-\alpha\tau} \propto ... \propto \tau^{n+g-1}e^{-\tau(\alpha+\sum x_i^2)}e^{-n\tau(\mu-\overline{X})^2}$, so $f(\mu|\tau, x)$ is $N(\bar{X}, n\tau)$ and $f(\tau|x)$ is $Gamma(n+g, \alpha+\sum x_i^2)$. The Bayes estimator is the posterior mean, $\frac{n+g}{\alpha+\sum x_i^2}$.

d) Under Squared Error Loss we need to consider the Bias & Variance of each estimator

For $\hat{\sigma}^2$. Bias is σ^2/n and variance $\frac{\sigma^4 2(n-1)}{n^2}$, so $MSE = \sigma^4\left(\frac{2n-1}{n^2}\right)$.

For S^2. Bias is 0 and variance is $\frac{2\sigma^4}{n-1}$, so $MSE = \frac{2\sigma^4}{n-1}$.

Since $\frac{2n-1}{n^2} < \frac{2}{n-1} = \frac{2n}{n^2-n}$, S^2 is inadmissible under Squared Error Loss.

Problem 4

a) Neymann-Pearson lemma.

b) Let $\theta_1 < \theta_2$. The Likelihood Ratio takes the form $\frac{\theta_2^x(1-\theta_2)^{n-x}}{\theta_1^x(1-\theta_1)^{n-x}}$, which is increasing with x. Hence a MP test will reject for large values of x.

$\phi(x) = \begin{cases} 1 & ,x > c \\ \beta & ,x = c \\ 0 & ,x < c \end{cases}$, where $P_{1/2}(X > c) + \beta P_{1/2}(X = c) = \alpha$.

Playing with different values we get $c = 3$ and $\beta = 0.14$.

c) By Karlin-Rubin the test in section b) is UMP.

d) Denote $\phi_1(X)$ the test defined in section b), which is the essentially unique UMP in the region $\theta \in (\frac{1}{2}, 1]$. Hence the test $\phi_1(X)$ is the only possible UMP, since no other test beats its in the region $\theta \in (\frac{1}{2}, 1]$. We'll show there's no UMP by seeing that $\phi_1(X)$ is no UMP.

Further the power $\gamma_1(\theta)$ of this test increases with θ (one of the corollaries of Karlin-Rubin gives this), and hence $\gamma_1(\theta) < \alpha$ for $\theta < 1/2$. But the test $\phi_2(X) = a$ has greater power and it's level α, and hence there's no UMP.

Problem 5

a) $\Rightarrow$) Let $Y = |X|$. $E(Y) = \int Y^p dP = \int_{\{Y\leq y\}} Y^p dP + \int_{\{Y>y\}} Y^p dP \geq \int_{\{Y\leq y\}} Y^p dP + y^p P(Y \geq y) \Rightarrow y^p P(Y \geq y) \leq E(Y) - \int_{\{Y\leq y\}} Y^p dP \rightarrow_{y\to\infty} 0$, as desired.

$\Leftarrow$) $E(|X|^{p-\epsilon}) = \int(p-\epsilon)|x|^{p-\epsilon-1}P(|X| > x)dx = \int_0^M (p-\epsilon)|x|^{p-\epsilon-1}P(|X| > x)dx + \int_M^\infty (p-\epsilon)|x|^{-(1+\epsilon)}|x|^p P(|X| > x)dx < \{|x|^p P(|X| > x) < N$ for large enough $N\} < \int_0^M (p-\epsilon)|x|^{p-\epsilon-1} P(|X|>x)dx + N\int_M^\infty (p-\epsilon)|x|^{-(1+\epsilon)}dx < \infty$, since the left term is the integral of a bounded function in a bounded interval, and the right term is finite since $|x|$ is raised to the negative of a number greater than one.

b) $\hat{\beta}_n = \bar{X}/3$. By Central Limit Theorem, $\sqrt{n}(\bar{X}-3\beta) \to N(0, 3\beta^2)$ and by Delta Method $\sqrt{n}(\bar{X}/3-\beta) \to N(0,\sigma^2)$, where $\sigma^2 = 3\beta^2\left(\frac{d}{d\beta}(\beta/3)\right)^2 = \beta^2/3$.

c) $P\left(-z_{\alpha/2} < \frac{\sqrt{n}(\bar{X}-3\beta)}{\sqrt{\hat{\beta}_n/3}} < z_{\alpha/2}\right) \to 1-\alpha$ by Slutsky and we obtain

$P\left(\bar{X}/3 - z_{\alpha/2}\sqrt{\hat{\beta}_n/3n} < \beta < \bar{X}/3 + z_{\alpha/2}\sqrt{\hat{\beta}_n/3n}\right) \to 1-\alpha.$

Problem 6

a) MGF of Gamma is $M_X(t) = (1-\beta t)^\alpha$ and hence $M_{X_1+X_2}(t) = (1-\beta t)^{\alpha_1+\alpha_2} \Rightarrow X_1+X_2 \sim Gamma(\alpha_1+\alpha_2, \beta)$

b) Let $\left\{\begin{array}{ll} Y_1 = X_1/X_2 & \Rightarrow X_1 = Y_1Y_2 \\ Y_2 = X_2 & \Rightarrow X_2 = Y_2 \end{array}\right.$. The Jacobian is $|J| = Y_2$ and hence $f_{Y_1Y_2}(y_1,y_2) = f_{X_1X_2}(y_1y_2, y_2)y_2andf_{Y_1}(y_1) = \frac{\Gamma(\alpha_1+\alpha_2-1)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta}\frac{y_1^{\alpha_1-1}}{(1+y_1)^{\alpha_1+\alpha_2-1}} I_{R^+}(y_1)$

Problem 7

a) Usual definition.

b) Proof of Basu's theorem.

c) Fix $\sigma^2 \in R^+$. Since $\frac{S^2}{\sigma^2}(n-1) \sim \chi^2_{n-1}$, $S^2$is ancillary for μ. Also we know that $\bar{X}$ is complete & sufficient for μ (check that the exponential family is full rank). Then, by Basu's theorem since $\bar{X}$ complete & sufficient for μ and S^2 ancillary for σ^2 we have that $\bar{X}$ and S^2 are independent.

Now, σ^2 was arbitrary so the result holds for all σ^2.

Problem 8

a) The Poisson belongs to the 1 parameter exponential family, so it has the MLR property. By Karlin Rubin the UMP test will reject for large values of the sufficient statistic, which is $\sum X_i$. We know that $\sum X_i \sim Poisson(n\lambda)$. Let k be the $max\{x : P_{\lambda_0}(\sum X_i < x) \leq \alpha\}$. Then the following test is UMP

$$\phi(X) = \begin{cases} 1 & , \sum X_i < k \\ \beta & , \sum X_i = k \\ 0 & , \sum X_i > k \end{cases}, \text{ where } \beta = \frac{\alpha - P_{\lambda_0}(\sum X_i < k)}{P_{\lambda_0}(\sum X_i = x)}$$

b) A conservative P-value would be $P_{\lambda_0}(\sum X_i < \sum x_i)$, where $\sum x_i$ is the observed value of the suff stat. A P-value is a test statistic with rejection region $[0, \alpha]$.

Statistics January 2004

Problem 1

a) MGF is $M_X(t) = (1-\beta t)^\alpha$, so $M_{X_1+X_2}(t) = (1-\beta t)^{\alpha_1+\alpha_2} \Rightarrow X_1 + X_2 \sim Gamma(\alpha_1 + \alpha_2, \beta)$

b) Let $\begin{cases} W_1 = \frac{X_1}{X_1+X_2} & \Rightarrow X_1 = W_1 W_2 \\ W_2 = X_1 + X_2 & \Rightarrow X_2 = W_2(1 - W_1) \end{cases}$, so the Jacobian is $|J| = w_2$ and $f_{w_1 w_2}(w_1, w_2) = f_{x_1 x_2}(w_1 w_2, w_2(1-w_1))w_2$. By integrating wrt w_2 we find that $f_{w_1}(w_1) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} w_1^{\alpha_1 - 1}(1-w_1)^{\alpha_2 - 1} I_{(0,1)}(w_1)$.

c) $P(R = r) = P(\frac{1}{Z} = r) = 2^{-1/r}$ for $r = 1, \frac{1}{2}, \frac{1}{3}...$

Problem 2

a) Use Cauchy-Schwartz inequality together with some regularity conditions.

b) $f_\theta(x) \propto \theta^x(1-\theta)^{n-x}$ so $\frac{d}{d\theta} log f_\theta(x) = 0 \Rightarrow \hat{\theta} = x/n$. Then $E\left(\frac{x}{\theta} - \frac{n-x}{1-\theta}\right)^2 = \frac{n}{\theta(1-\theta)}$ and hence $V_\theta(T(X)) \geq \frac{\theta(1-\theta)}{n}$ and $V_\theta(\bar{X}) = \frac{V(X)}{n^2} = \frac{\theta(1-\theta)}{n}$ attains the Cramer-Rao LB.

Problem 3

a) $\bar{X} = \frac{1}{n}\sum X_i$ is a linear combination of independent & normal rv's and hence it's normally distributed with mean μ and variance $1/n$. Equivalently $\frac{\bar{X}-\mu}{1/\sqrt{n}} \sim N(0,1)$, and the rejection region is given by finding z_1, z_2 such that $P_{H_0}(z_1 < Z < z_2) = 1-\alpha$, e.g. $z_1 = -z_{\alpha/2}$ and $z_2 = z_{\alpha/2}$.

b) $\gamma(\mu) = 1 - P_{H_0}(-z_{\alpha/2} < Z < z_{\alpha/2})$. For general μ, $\bar{X} \sim N(\mu, 1/n) \Rightarrow Z = \sqrt{n}(\bar{X}-10) \sim N(\sqrt{n}(\mu-10), 1)$, and hence we can use the quantiles of this latter normal distribution to get the desired result.

c) Not UMP because the UMP for $\mu > 10$ rejects for large values of X (by Karlin-Rubin) and for $\mu < 10$ rejects for small values of X, i.e. the UMP test for $\mu > 10$ is different than the UMP test for $\mu < 10$ and since they're both unique there can be no UMP test.

d) LR test rejects for large values of $\frac{e^{-\frac{1}{2}\sum(x_i-\bar{X})^2}}{e^{-\frac{1}{2}\sum(x_i-10)^2}}$ or equivalently for large $\sum(x_i-10)^2 - \sum(x_i-\bar{X})^2$. With some algebra we can see that this is equivalent to rejecting for large $|Z|$. Rejecting for $|Z| > z_{\alpha/2}$gives a level α test.

Problem 4

Times series, not necessary.

Problem 5

a) Factorization theorem.

b) By Rao-Blackwell any unbiased estimator can be improved by finding its exectation conditional on the suff stat, and so it follows that the best unbiased estimator has to be a function of the suff . Note that the UMVU may not be unique if we don't have completeness.

Now, noting that the likelihood will be of the form $log\,(g(T(x),\theta) + log(h(x))$, we can drop the second term since it doesn't depend on θ and maximize $log\,(g(T(x),\theta))$ which will depend on the data only through $T(x)$, the sufficient stat.

The Bayes estimator is of the form $E(\theta|X) = \int \theta f(\theta|X) d\theta$ and $f(\theta|X) = \frac{g(T(x),\theta)h(x)\pi(\theta)}{\int g(T(x),\theta)h(x)\pi(\theta)d\theta} = \frac{g(T(x),\theta)\pi(\theta)}{\int g(T(x),\theta)\pi(\theta)d\theta}$ which depends on the data only through $T(x)$.

c) $lim_{n\to\infty} E_\theta(\delta_n(x)) = \theta$, for all $\theta \in \Theta$.

d) (i) Likelihood is $f_\mu(x) = (2\pi)^{-n/2} e^{-\frac{1}{2}\sum(x_i-\mu_i)^2}$, and taking log and then derivative and setting equal to zero gives $x_i = \mu_i$. Hence $(\hat{\mu}_1 ... \hat{\mu}_n) = (x_1 ... x_n)$ and by the invariance property of MLEs we get $\hat{\theta}_n = \frac{1}{n}\sum \hat{\mu}_i^2 = \frac{1}{n}\sum x_i^2$.

(ii) We can find $E(X_j^2) = \mu_j^2 + 1$ by using the 2nd derivative of the characteristic function of X_j. Hence $E(\hat{\theta}_n) = 1 + \frac{1}{n}\sum \mu_j^2$ and thus θ_n^* is unbiased. Now, we know that $(x_1 ... x_n)$ is a sufficient statistic and it can be seen that the exponential family is full rank, so the sufficient statistic is also complete. Hence Lehman-Scheffe applies and θ_n^* is UMVU. Also, $V(\theta_n^*) = V(\hat{\theta}_n - 1) = V(\hat{\theta}_n)$ and $Bias(\theta_n^*) < Bias(\hat{\theta}_n)$, so $\hat{\theta}_n$ is inadmissible under squared error loss.

(iii) If we can show that $\theta_n^* \to^P \theta$, then $\hat{\theta}_n \to \theta + 1$ by the Continuous Mapping Principle. I'm not sure how to do it.

Problem 6

Easy but long!

Problem 7

a) The original model is $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2 \rho_i)$ independent, so $E(y_i/\rho_i) = \beta_0/\rho_i + \beta_1 x_i \rho_i + E(\epsilon_i/\rho_i) \Rightarrow E(\delta_i) = E(\epsilon_i/\rho_i) = 0$. Hence δ_i are independent with zero means and variance σ^2/ρ_i.

b) The usual least squares estimators give $\hat{\beta}_1 = \frac{\sum(v_i-\bar{v})(z_i-\bar{z})}{\sum(v_i-\bar{v})^2}$ where $\bar{v}$ and $\bar{z}$ are the sample means, and $\hat{\beta}_0 = \bar{z} - \hat{\beta}_1\bar{v}$. With some algebra we can see that they're not unbiased.

c) and d) Long!

Problem 8

Biostatistics.