Mathematical Statistics: Ph.D. Qualifier — October 4, 2001

1. (20pts). Let $\{X_n\}_{n=1}^{\infty}$ be Bernoulli random variables with the following joint distribution P_θ:

$$P_\theta(X_1 = 1) = \theta_1$$

$$P_\theta(X_i = 1 | X_1, \cdots, X_{i-1}) = \begin{cases} \theta_{11} & \text{if } X_{i-1} = 1, \\ \theta_{01} & \text{if } X_{i-1} = 0, \end{cases} \quad i = 2, 3, \cdots,$$

where $\theta = (\theta_1, \theta_{11}, \theta_{01})$. Let $X = (X_1, \ldots, X_n), n \geq 3$. Find a four-dimensional sufficient statistics.

2. Suppose that $\{X_n\}_{n=1}^{\infty}$ are conditionally IID given θ with conditional density

$$f_\alpha(x_i|\theta) = \frac{\theta^\alpha \alpha}{x_i^{\alpha+1}} I_{[\theta,\infty)}(x_i),$$

$\alpha > 0$, $\theta > 0$

where α is a known positive number and the parameter space is $\theta \in \Theta = (0, \infty)$.

   (a) (15pts). Find the MLE $\hat{\theta}$ of θ based on $X_1, \ldots, X_n$.

   (b) (20pts). Show that the MLE is inadmissible if the loss is squared error. (Assume $n\alpha > 2$.)

   (c) (20pts). Find a_n and b_n such that $a_n\hat{\theta} + b_n$ has a nondegenerate asymptotic distribution, and find that distribution.

3. (25pts). For each $k = 1, 2, \ldots$, let $\gamma_{k,\alpha}$ denote the $1-\alpha$ quantile of the χ_k^2 distribution. Let Y_k have a noncentral χ_k^2 distribution with noncentrality parameter c^2. Prove that $P(Y_k > r_{k,\alpha}) \leq P(Y_1 > r_{1,\alpha})$ for all $k \geq 1$ and $c^2 \geq 0$.
(Hint: Let $X_1, \ldots, X_k$ be IID $N(\theta, 1)$. Consider two tests of $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$ based on $\bar{X}^2$ and $\Sigma_{i=1}^k X_i^2$. Note that one of the statistics gives us a UMPU test.)

(2)

1. Suppose that $Y_1, \ldots, Y_n$ are iid with pdf 20 (~~15~~%)

$$\frac{1}{2\sigma}\exp\{-|y-\mu|/\sigma\}, \quad -\infty < y < \infty .$$

Consider the problem of testing the hypothesis $\mu = 0$ against the one-sided alternative $\mu > 0$.

(i) Show that, under H_0, there is a boundedly complete sufficient statistic S for σ.

(ii) What does this imply about the size of a similar test conditional on S = s ?

(iii) Show that under the null hypothesis S has a gamma distribution, and hence that the density of Y|S=s is a uniform on the set S = s.

(iv) Hence show that the likelihood ratio is a monotone function of $\sum|Y_i - \mu|$, and that a uniformly most powerful similar test does not exist.

(v) Show that the locally most powerful similar test is based on $\sum$ sign Y_i .

2. Suppose that $Y_1, \ldots, Y_n$ are i.i.d. N(μ, σ^2), and that we wish to test the hypothesis $\sigma = \sigma_0$ against the alternative $\sigma > \sigma_0$, the mean μ being unrestricted in both cases ($-\infty < \mu < \infty$) . (16%)

(a) Find minimal sufficient statistics under both hypotheses.

(b) Show that both hypotheses are invariant under the location shift group G = { $g_a ; -\infty < a < \infty$ }, where g_a is defined by

$$g_a : (y_1, \ldots, y_n) \rightarrow (y_1\text{+a}, \ldots, y_n\text{+a}) .$$

(c) Find the maximal invariant function of the statistic that is minimal sufficient under the alternative hypothesis.

(d) Without carrying out any further calculations, state how you would obtain the uniformly most powerful invariant test, and give the form of that test.

3. A testing laboratory has n technicians, $T_1, \ldots, T_n$. When an item characterized by the value μ is measured by the i^{th} technician, T_i, the resulting (single) measurement is distributed as N(μ, σ_i^2), where the variance $\sigma_1^2, \ldots, \sigma_n^2$ are all known. Items are assigned to technicians randomly, with equal probabilities. (14%)

(a) Write down the joint distribution of (Y, T), where Y is the measurement and T is the identity of the technician who made it.

(b) Show that T is ancillary.

(c) Find the most powerful test of the hypothesis $\mu = \mu_0$ against the alternative $\mu =$

$\mu_0+\delta$ for a fixed $\delta > 0$.

(d) Repeat part (c) based on the conditional distribution of Y|T. What principle asserts that the latter test is more appropriate, even though less powerful?

4. Consider the bivariate disctete distribution with density

$$P(x, y; \alpha, \beta, \gamma) = c\frac{\alpha^x \beta^y \gamma^{xy}}{x!\, y!}\ ;\ x=0,1,2,\ldots\ ;\ y=0,1,2,\ldots$$

where $\alpha \geq 0, \beta \geq 0, \gamma \geq 0$ and c is a normalizing constant. (25%)

(a) Find the marginal discrete density $P_1(x;\alpha,\beta,\gamma)$,and show that

$\sum_x p_1(x;\alpha,\beta,\gamma)$ is not finite unless $\gamma \geq 1$.

So what is the natural parameter space in terms of α,β,γ?

(b) Determine the conditional distribution of Y given X ; show it is Poisson!

(c) Find the regression of Y on X ; what does this regression line look like?

(d) Which value of γ corresponds to X and Y being independent?

Now consider a random sample of size n from the above distribution:

$(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$.

The purpose of the remaining questions is the construction of the UMP test of appropriate size of the null hypothesis that X and Y are independent (in the above distribution from which the sample was taken)

vs. the alternative that X and Y are dependent.

(e) Why may we claim that a UMP test exists rather than a UMPU test?

(f) Given the minimal sufficient statistic for the entire family ; remember to prove affine independent of its components.

(g) What is the conditioning statistic for this problem?

(h) In order to implement this test , one needs to make cumbersome calculations starting with the enumeration of possible point in sample space . Do these calculation for n=2 , $(X_1,Y_1)=(2,3)$ and $(X_2,Y_{21})=(1,1)$.

5. A. Let X and Y be independent random variables with continuous probability density function f(x) and g(y) , respectively. Show that (~~20~~ 25%)

$$\Pr\{X \geqq Y\} = \int_{\infty}^{\infty} G(x)f(x)dx = \int_{\infty}^{+\infty}[1-F(y)]g(y)dy \quad ,$$

Where F and G are distribution function of X and Y, respectively. Hence, show that if f=g, then the Pr(X≧Y)=1/2.

B. Let $Y_1, Y_2, \ldots Y_{n+1}$ be (n+1) independent normal random variables with mean 0 and variance $\sigma^2/2$.

(a) Show that if $X_j = Y_j - Y_{j+1}$, j=1,…,n, then $E(X_j)=0$, $Var(X_j)=\sigma^2$,

and $Cov(X_j, X_k)=\sigma^2/2$, j≠k.

(b) Show that $P(X_1 \geq 0, \ldots, X_n \geq 0) = \dfrac{1}{n+1}$.

(c) Let $Z_j = Y_j - Y_{j+1}$, j=1,…,n. Find $Z(X_j)$, $Var(Z_j)$, and $Cov(Z_j, Z_k)$, j≠ k.

(d) Show that $P(Z_1 \geq 0, \ldots, Z_n \geq 0) = \dfrac{1}{(n+1)!}$.

③

PhD QR exam 90 MathStat

1. Let $X_1, ..., X_n$ be iid with distribution $F(x-\theta)$ and density f. Suppose that $F(0) = 1/2$ and $f(0) > 0$. Denote the sample mean and sample median as $\bar{X}_n$ and $\tilde{X}_n$, respectively.
   (a) (20 points) Show that $\sqrt{n}(\tilde{X}_n - \theta) \Rightarrow N(0, 1/[4f^2(0)])$.
   (b) (10 points) Let $\sigma^2 = Var(X_1)$. Show that the asymptotic relative efficiency (ARE) of $\tilde{X}_n$ to $\bar{X}_n$ is
   $$e_{\tilde{X}_n, \bar{X}_n}(F) = 4f^2(0)\sigma^2.$$
   (c) (10 points) Suppose that
   $$F(x) = (1-\epsilon)\Phi(x) + \epsilon\Phi(\frac{x}{\tau}).$$
   Calculate $e_{\tilde{X}_n, \bar{X}_n}(F)$ for $\epsilon = 0.01, 0.1$ and $\tau = 2, 4$.

2. Let $X_1, ..., X_n$ be iid from $N(0, \tau)$ population.
   (a) (10 points) Give the meaning of "conjugate prior."
   (b) (15 points) What is the conjugate prior of τ? (Verify your answer.)
   (c) (15 points) Find Bayes estimator of τ using conjugate prior and loss function $L(\tau, d) = (\tau - d)^2/\tau^2$.

3. (20 points) Assume the probabilistic model:
   $$p(x|\Theta) = \sum_{i=1}^{M} \alpha_i p_i(x|\theta_i),$$
   where the parameters are $\Theta = (\alpha_1, \cdots, \alpha_M, \theta_1, \cdots, \theta_M)$ such that $\sum_{i=1}^{M} \alpha_i = 1$ and $\theta_i = (\mu_i, \sigma_i^2)$, and each p_i is a normal density function with mean μ_i and variance σ_i^2. In other words, we assume that we have M component normal densities mixed together with M mixing coefficients α_i. Let $X = (X_1, \cdots, X_N)'$ be the observed data and $Y = (Y_1, \cdots, Y_N)'$ be the unobserved data with $Y_i \in 1, \cdots, M$ for each i, and $Y_i = k$ if X_i is generated by the kth mixture component.

   Define $Q(\Theta, \Theta')$ as the conditional expectation of $\log p(X, Y|\Theta)$ given X and the current parameter estimates Θ'. That is,
   $$Q(\Theta, \Theta') = E\Big[\log p(X, Y|\Theta)|X, \Theta'\Big].$$
   Find $\text{argmax}_\Theta Q(\Theta, \Theta')$.

(2)

1. Suppose that $Y_1, \ldots, Y_n$ are iid with pdf (~~15~~ 20%)

$$\frac{1}{2\sigma}\exp\{-|y-\mu|/\sigma\}, \quad -\infty < y < \infty .$$

Consider the problem of testing the hypothesis $\mu = 0$ against the one-sided alternative $\mu > 0$.

(i) Show that, under H_0, there is a boundedly complete sufficient statistic S for σ.

(ii) What does this imply about the size of a similar test conditional on S = s ?

(iii) Show that under the null hypothesis S has a gamma distribution, and hence that the density of Y|S=s is a uniform on the set S = s.

(iv) Hence show that the likelihood ratio is a monotone function of $\sum|Y_i - \mu|$, and that a uniformly most powerful similar test does not exist.

(v) Show that the locally most powerful similar test is based on $\sum \operatorname{sign} Y_i$.

2. Suppose that $Y_1, \ldots, Y_n$ are i.i.d. N(μ, σ^2), and that we wish to test the hypothesis $\sigma = \sigma_0$ against the alternative $\sigma > \sigma_0$, the mean μ being unrestricted in both cases ($-\infty < \mu < \infty$). (16%)

(a) Find minimal sufficient statistics under both hypotheses.

(b) Show that both hypotheses are invariant under the location shift group G = { g_a ; $-\infty < a < \infty$ }, where g_a is defined by

$$g_a : (y_1, \ldots, y_n) \rightarrow (y_1\text{+a}, \ldots, y_n\text{+a}) .$$

(c) Find the maximal invariant function of the statistic that is minimal sufficient under the alternative hypothesis.

(d) Without carrying out any further calculations, state how you would obtain the uniformly most powerful invariant test, and give the form of that test.

3. A testing laboratory has n technicians, $T_1, \ldots, T_n$. When an item characterized by the value μ is measured by the i^{th} technician, T_i, the resulting (single) measurement is distributed as N(μ, σ_i^2), where the variance $\sigma_1^2, \ldots, \sigma_n^2$ are all known. Items are assigned to technicians randomly, with equal probabilities. (14%)

(a) Write down the joint distribution of (Y, T), where Y is the measurement and T is the identity of the technician who made it.

(b) Show that T is ancillary.

(c) Find the most powerful test of the hypothesis $\mu = \mu_0$ against the alternative $\mu =$

$\mu_0+\delta$ for a fixed $\delta>0$.

(d) Repeat part (c) based on the conditional distribution of Y|T . What principle asserts that the latter test is more appropriate, even though less powerful?

4. Consider the bivariate disctete distribution with density

$$P(x,y;\alpha,\beta,\gamma)=c\,\frac{\alpha^x\beta^y\gamma^{xy}}{x!\,y!}\ ;\ x=0,1,2,\ldots\ ;\ y=0,1,2,\ldots$$

where $\alpha\ge 0,\beta\ge 0,\gamma\ge 0$ and c is a normalizing constant. (25%)

(a) Find the marginal discrete density $P_1(x;\alpha,\beta,\gamma)$,and show that

$\sum_x p_1(x;\alpha,\beta,\gamma)$ is not finite unless $\gamma\ge 1$.

So what is the natural parameter space in terms of α,β,γ ?

(b) Determine the conditional distribution of Y given X ; show it is Poisson!

(c) Find the regression of Y on X ; what does this regression line look like?

(d) Which value of γ corresponds to X and Y being independent?

Now consider a random sample of size n from the above distribution:

$(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$.

The purpose of the remaining questions is the construction of the UMP test of appropriate size of the null hypothesis that X and Y are independent (in the above distribution from which the sample was taken)

vs. the alternative that X and Y are dependent.

(e) Why may we claim that a UMP test exists rather than a UMPU test?

(f) Given the minimal sufficient statistic for the entire family ; remember to prove affine independent of its components.

(g) What is the conditioning statistic for this problem?

(h) In order to implement this test , one needs to make cumbersome calculations starting with the enumeration of possible point in sample space . Do these calculation for n=2 , $(X_1,Y_1)=(2,3)$ and $(X_2,Y_{21})=(1,1)$.

5. A. Let X and Y be independent random variables with continuous probability density function f(x) and g(y) , respectively. Show that (~~20~~ 25%)

$$\Pr\{X\geqq Y\}=\int_{-\infty}^{+\infty}G(x)f(x)dx=\int_{-\infty}^{+\infty}[1-F(y)]g(y)dy\ ,$$

Where F and G are distribution function of X and Y, respectively. Hence, show that if f=g, then the Pr(X≧Y)=1/2.

B. Let $Y_1, Y_2, \ldots Y_{n+1}$ be (n+1) independent normal random variables with mean 0 and variance $\sigma^2/2$.

(a) Show that if $X_j = Y_j - Y_{j+1}$, j=1,…,n, then $E(X_j)=0$, $Var(X_j)=\sigma^2$,

and $Cov(X_j, X_k)=\sigma^2/2$, j≠k.

(b) Show that $P(X_1 \geq 0, \ldots, X_n \geq 0) = \dfrac{1}{n+1}$.

(c) Let $Z_j = Y_j - Y_{j+1}$, j=1,…,n. Find $Z(X_j)$, $Var(Z_j)$, and $Cov(Z_j, Z_k)$, j≠ k.

(d) Show that $P(Z_1 \geq 0, \ldots, Z_n \geq 0) = \dfrac{1}{(n+1)!}$.

Ph.D. 資格考 10月 90年.

PhD QR exam 90 MathStat

1. Let $X_1, ..., X_n$ be iid with distribution $F(x-\theta)$ and density f. Suppose that $F(0) = 1/2$ and $f(0) > 0$. Denote the sample mean and sample median as $\bar{X}_n$ and $\tilde{X}_n$, respectively.
(a) (20 points) Show that $\sqrt{n}(\tilde{X}_n - \theta) \Rightarrow N(0, 1/[4f^2(0)])$.
(b) (10 points) Let $\sigma^2 = Var(X_1)$. Show that the asymptotic relative efficiency (ARE) of $\tilde{X}_n$ to $\bar{X}_n$ is
$$e_{\tilde{X}_n, \bar{X}_n}(F) = 4f^2(0)\sigma^2.$$
(c) (10 points) Suppose that
$$F(x) = (1-\epsilon)\Phi(x) + \epsilon\Phi(\frac{x}{\tau}).$$
Calculate $e_{\tilde{X}_n, \bar{X}_n}(F)$ for $\epsilon = 0.01, 0.1$ and $\tau = 2, 4$.

2. Let $X_1, ..., X_n$ be iid from $N(0, \tau)$ population.
(a) (10 points) Give the meaning of "conjugate prior."
(b) (15 points) What is the conjugate prior of τ? (Verify your answer.)
(c) (15 points) Find Bayes estimator of τ using conjugate prior and loss function $L(\tau, d) = (\tau - d)^2/\tau^2$.

3. (20 points) Let $x = (x_1, \cdots, x_N)'$. Assume the probabilistic model:
$$p(x|\Theta) = \sum_{i=1}^{M} \alpha_i p_i(x|\theta_i),$$
where the parameters are $\Theta = (\alpha_1, \cdots, \alpha_M, \theta_1, \cdots, \theta_M)$ such that $\sum_{i=1}^{M} \alpha_i = 1$ and $\theta_i = (\mu_i, \sigma_i^2)$, and each p_i is a normal density function with mean μ_i and variance σ_i^2. In other words, we assume that we have M component normal densities mixed together with M mixing coefficients α_i. Let $y = (y_1, \cdots, y_N)'$ be the unobserved data with $y_i \in 1, \cdots, M$ for each i, and $y_i = k$ if x_i is generated by the kth mixture component.

Define $Q(\Theta, \Theta')$ as the conditional expectation of $\log p(x, y|\Theta)$ given x and the current parameter estimates Θ'. That is,
$$Q(\Theta, \Theta') = E\Big[\log p(x, y|\Theta)|x, \Theta'\Big].$$
Find $\text{argmax}_\Theta Q(\Theta, \Theta')$.