Statistics 583, Problem Set 5 Solutions

Wellner; 4/26/2000

1. Let X, Y be independent exponential random variables with parameters μ and ν respectively:
$$p_{\mu,\nu}(x,y) = \mu\nu\exp(-\mu x - \nu y)1_{(0,\infty)}(x)1_{(0,\infty)}(y)\,.$$
Let $\theta = \mu/\nu$. Show that the problem of testing $H : \theta \le 1$ versus $K : \theta > 1$ is invariant under the group of transformations $g_c(x,y) = (cx, cy)$, $c > 0$, and find a UMP invariant test of size α.

Solution: For $c > 0$, the distribution of cX is exponential(μ/c):
$$P_\mu(cX > x) = P_\mu(X > x/c) = \exp(-\mu x/c) = \exp(-(\mu/c)x)\,.$$
Similarly, the distribution of cY is exponential(ν/c). Hence the induced group $\overline{G}$ on the parameter space $\Theta = R^{+2}$ is $\overline{g}_c(\mu,\nu) = (\mu/c, \nu/c)$. The null hypothesis $\Theta_0 = \{(\mu,\nu) : \mu/\nu \le 1\}$ is clearly invariant: $\overline{g}_c\Theta_0 = \Theta_0$; and Θ is also invariant, $\overline{g}_c\Theta = \Theta$. It is easily seen that $T(Y/X)$ is a $G-$maximal invariant: T is invariant since $T(cX, cY) = cY/(cX) = Y/X = T(X,Y)$, and T is maximal since $T(X^*, Y^*) = T(X,Y)$ implies $Y^*/X^* = Y/X$, or $Y^* = Y(X^*/X) = cY$ with $c \equiv X^*/X$, and then $(X^*, Y^*) = (cX, cY) = g_c(X,Y)$, so that (X,Y) and (X^*,Y^*) are on the same orbit. Furthermore, $\theta = \mu/nu$ is a $\overline{G}-$maximal invariant. Note that the maximal invariant T can be written as
$$\begin{aligned} T(X,Y) &= \left(\frac{\nu Y}{\mu X}\right)\left(\frac{\mu}{\nu}\right) = \left(\frac{\nu Y/2}{\mu X/2}\right)\left(\frac{\mu}{\nu}\right) \\ &=_d \frac{\chi_2^2/2}{\chi_2^2/2}\left(\frac{\mu}{\nu}\right) =_d \theta F_{2,2}\,. \end{aligned}$$
Since the density of an $F_{2,2}$ variable is $f(x) = (1+x)^{-2}1_{(0,\infty)}(x)$, the density of $\theta F_{2,2}$ is $\theta^{-1}f(x/\theta)$, and it is easily seen that this has monotone likelihood ratio in x: for $\theta_1 < \theta_2$,
$$\frac{\theta_2^{-1}f(x/\theta_2)}{\theta_1^{-1}f(x/\theta_1)} = \frac{\theta_1}{\theta_2}\frac{(1+x/\theta_1)^2}{(1+x/\theta_2)^2}$$
which is increasing in x. Thus the hypotheses of Theorem 6.3.4.2 are satisfied, and the UMP G-invariant test of H versus K under the group G is given by
$$\phi(X,Y) = \begin{cases} 1 & \text{if} \quad Y/X > c \\ \gamma & \text{if} \quad Y/X = c \\ 0 & \text{if} \quad Y/X < c\,. \end{cases}$$
where c and γ are chosen so that $E_{\mu,\mu}\phi(X,Y) = \alpha$. But under the boundary of the null hypothesis we have $Y/X \sim F_{2,2}$, and $P(F_{2,2} > c) = 1/(1+c)$, and this equals α if $c = (1-\alpha)/\alpha$. Thus we can take $\gamma = 0$, and the UMP G-invariant test becomes "reject H_0 if $T > (1-\alpha)/\alpha$".

2. What happens in the context of problem 1 if we observe $X_1, \ldots, X_m$ i.i.d. as X and $Y_1, \ldots, Y_n$ i.i.d. as Y? Can you find a UMP invariant test of H versus K?

Solution: When $X_1, \ldots, X_m$ are i.i.d. Exponential(μ), and $Y_1, \ldots, Y_n$ are i.i.d. Exponential(ν), the sufficient statistics are $R \equiv \sum_1^m X_i$ and $S \equiv \sum_1^n Y_j$, where $R \sim$ Gamma(m, μ) and $S \sim$ Gamma(n, ν). The distributions of cR and cS are Gamma$(m, \mu/c)$ and Gamma$(n, \nu/c)$ respectively, so the testing problem continues to be invariant under the group G and the induced group G^* on the space of the sufficient statistics, R^{+2}. A maximal invariant on the space of the sufficient statistics is

$$
\begin{aligned}
T(R,S) &= \frac{m}{n}\frac{S}{R} = \frac{m}{n}\frac{\nu S}{\mu R}\frac{\mu}{\nu} \\
&= \frac{\mu}{\nu}\frac{2\nu S/2n}{2\mu R/2m} \\
&= \theta \frac{\chi^2_{2n}/2n}{\chi^2_{2m}/2m} \sim \theta F_{2n,2m} .
\end{aligned}
$$

The density of $F_{2n,2m}$ is

$$
f_{2n,2m}(x) = \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} \left(\frac{n}{m}\right)^n \frac{x^{n-1}}{(1+(n/m)x)^{n+m}} .
$$

Thus the density of $\theta F_{2n,2m}$ is $\theta^{-1} f_{2n,2m} x/\theta)$, and this has monotone likelihood ratio in x: for $\theta_1 < \theta_2$,

$$
\frac{\theta_2^{-1} f_{2n,2m}(x/\theta_2)}{\theta_1^{-1} f_{2n,2m}(x/\theta_1)} = \left(\frac{\theta_1}{\theta_2}\right)^n \frac{(1+(n/m)x/\theta_1)^{n+m}}{(1+(n/m)x/\theta_2)^{n+m}}
$$

which is increasing in x. Thus the hypotheses of Theorem 6.3.4.2 are satisfied, and the UMP G-invariant test of H versus K is given by "reject H if $T(R,S) > F_{2n,2m,\alpha}$".

3. Suppose that an urn contains N balls with the numbers $z_i = -\log(1 - i/(N+1))$, $i = 1, \ldots, N$ and we sample $n < N$ balls from this urn. Let $\overline{Y}_n = n^{-1}\sum_1^n Y_i$ denote the sample mean of the sampled valls.
A. Calculate the mean $\mu_N = E(\overline{Y}_n)$ and variance $\sigma_N^2 = Var(\overline{Y}_n)$ of $\overline{Y}_n$.
Find the limits of $\overline{z}_N$ and σ_z^2 as $N \to \infty$.
B. Use the Wald-Wolfowitz-Noether-Hajek finite-sampling CLT to prove that $(\overline{Y}_n - \mu_N)/\sigma_N \to_d N(0,1)$.
C. What classical two-sample rank statistic is $\overline{Y}_n$ equivalent to under the null hypothesis (of all $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ equal in distribution with a common continuous distribution function F)?

Solution: A. The mean is

$$
\begin{aligned}
\mu_N &= E(\overline{Y}_n) = \overline{z}_N \\
&= \frac{1}{N}\sum_{i=1}^N \left\{ -\log\left(1 - \frac{i}{N+1}\right)\right\} \\
&\to \int_0^1 \{-\log(1-t)\}\, dt = 1
\end{aligned}
$$

upon noticing that $F^{-1}(t) = -\log(1-t)$ for the standard exponential distribution $F(x) = 1 - e^{-x}$, $x \geq 0$, so that $F^{-1}(U) =_d Y \sim$ Exponential(1). Similarly, the variance is

$$\sigma_N^2 = Var(\overline{Y}_n) = \frac{\sigma_z^2}{n}\left(1 - \frac{n-1}{N-1}\right),$$

where

$$\begin{aligned} \sigma_z^2 &= \frac{1}{N}\sum_{i=1}^{N}(z_i - \overline{z}_N)^2 \\ &\to \int_0^1 \{-\log(1-t) - 1\}^2 dt \\ &= Var(Y) = 1\,. \end{aligned}$$

B. The Wald-Wolfowitz-Noether-Hájek finite-sampling CLT yields $(\overline{Y}_n - \mu_N)/\sigma_N \to_d N(0,1)$ as long as $0 < \liminf(n/N) \leq \limsup(n/N) < 1$ if we show that the Noether condition holds. But the Noether condition is

$$\eta_N \equiv \frac{\max_{1\leq i\leq N}|z_i - \overline{z}_N|}{\sum_1^N (z_i - \overline{z}_N)^2} \to 0\,.$$

Upon dividing the numerator and denominator by N, we know from part A that the denominator (divided by N) converges to 1. Hence it suffices to show that

$$N^{-1}\max_{1\leq i\leq N}|z_i - \overline{z}_N|^2 \to 0\,.$$

Now since z_i increases with i,

$$\begin{aligned} \max_{1\leq i\leq N}|z_i - \overline{z}_N| &\leq \max_{1\leq i\leq N}(\overline{z}_N - z_i) \vee \max_{1\leq i\leq N}(z_i - \overline{z}_N) \\ &\leq \overline{z} \vee (z_N - \overline{z}_N) \end{aligned}$$

where $z_N = -\log(1 - N/(N+1)) = -\log(1/(N+1)) = \log(N+1)$. Thus we have

$$\begin{aligned} &N^{-1}\max_{1\leq i\leq N}|z_i - \overline{z}_N|^2 \\ &\leq N^{-1}\overline{z}_N^2 \vee N^{-1}(\log(N+1) - \overline{z}_N)^2 \\ &\to 0 \vee 0 = 0\,. \end{aligned}$$

C. Under the null hypothesis $\overline{Y}_n$ is equivalent to the "log-rank" statistic

$$T_N \equiv \frac{1}{n}\sum_{i=1}^{n}\left\{-\log\left(1 - \frac{R_i}{N+1}\right)\right\}$$

where R_i is the rank of Y_i, $i = 1, \ldots, n$ in the combined sample, $X_1, \ldots, X_m, Y_1, \ldots, Y_n$.