Topics in Statistical Theory — Example Sheet 1

Lucas Riedstra

...

Question 1. *Let $U_1, \ldots, U_n \overset{\text{iid}}{\sim} U(0,1)$ and let $Y_1, \ldots, Y_{n+1} \overset{\text{iid}}{\sim} \text{Exp}(1)$. Writing $S_j \coloneqq \sum_{i=1}^j Y_i$ for $j = 1, \ldots, n+1$, show that*

$$U_{(j)} \overset{\text{d}}{=} \frac{S_j}{S_{n+1}} \sim \text{Beta}(j, n-j+1)$$

for $j = 1, \ldots, n$.

Solution. First we compute the distribution function of $X_{(j)}$. If $X_{(j)} \le x$, then at least j of the X_i must be $\le x$.

For $k = j, j+1, \ldots, n$, there are $\binom{n}{k}$ ways to choose k of the X_i's that must be $\le j$. Fix $x \in [0,1]$, then we have

$$\begin{aligned}
\mathbb{P}(X_{(j)} \le x) &= \sum_{k=j}^{n} \binom{n}{k} \mathbb{P}(X_1 \le x, \ldots, X_k \le x, X_{k+1} > x, \ldots, X_n > x) \\
&= \sum_{k=j}^{n} \binom{n}{k} (F(x))^k \cdot (1 - F(x))^{n-k} \\
&= \sum_{k=j}^{n} \binom{n}{k} x^k (1-x)^{n-k}.
\end{aligned}$$

Now, the density function of $X_{(k)}$ is given by the derivative of F. This is a hefty calculation, see for example this StackExchange link, but it turns out to be

$$\frac{\mathrm{d}}{\mathrm{d}x} F_{(j)}(x) = f_{(j)}(x) = \frac{n!}{(j-1)!(n-j)!} x^{j-1} (1-x)^{n-j},$$

which is also the density function of the Beta$(j, n-j+1)$ distribution.

Finally, define $T = S_{n+1} - S_j$, so that S_j and T are independent. It is known that $S_j \sim \text{Gamma}(j, 1)$, $T \sim \gamma(n-j+1)$, and furthermore that

$$\frac{S_j}{S_{n+1}} = \frac{S_j}{S_j + T} \overset{\text{d}}{=} \frac{\Gamma(j,1)}{\Gamma(j,1) + \Gamma(n-j+1,1)} \sim \text{Beta}(j, n-j+1).$$

Question 2. *Let X be a random variable with mean zero that satisfies $a \le X \le b$. Use convexity to show that for every $t \in \mathbb{R}$, we have*

$$\log \mathbb{E}(e^{tX}) \le -\alpha u + \log(\beta + \alpha e^u),$$

where $u \coloneqq t(b-a)$ and $\alpha \coloneqq 1 - \beta = -a/(b-a)$. Using a second-order Taylor expansion around the origin, deduce that $\log \mathbb{E}(e^{tX}) \le t^2(b-a)^2/8$.

Proof. Let $x \in [a, b]$, then we know there exists a unique $\lambda \in [0, 1]$ such that $x = (1 - \lambda)a + \lambda b$. A simple computation gives $\lambda = (x - a)/(b - a)$, $1 - \lambda = (b - x)/(b - a)$. By convexity of $t \mapsto e^{tx}$ we find

$$e^{tx} \leq \frac{b-x}{b-a} e^{ta} + \frac{x-a}{b-a} e^{tb}.$$

From this we deduce that

$$\mathbb{E}[e^{tX}] \leq \mathbb{E}\left[\frac{b-X}{b-a} e^{ta} + \frac{X-a}{b-a} e^{tb}\right] = \frac{b}{b-a} e^{ta} + \frac{-a}{b-a} e^{tb} = \beta e^{ta} + \alpha e^{tb} = e^{-\alpha u + \log(\beta + \alpha e^u)}.$$

Since log is increasing, we can take the logarithm on both sides to conclude

$$\log \mathbb{E}[e^{tX}] \leq -\alpha u + \log(\beta + \alpha e^u).$$

Now, we compute the taylor polynomial of $f(u) \coloneqq -\alpha u + \log(\beta + \alpha e^u)$ in $u = 0$: we have

$$\begin{aligned}
f(0) &= \log(\beta + \alpha) = \log(1) = 0; \\
f'(u) &= -\alpha + \frac{\alpha e^u}{\beta + \alpha e^u}; \\
f'(0) &= -\alpha + \frac{\alpha}{\beta + \alpha} = 0; \\
f''(u) &= \frac{(\beta + \alpha e^u)\alpha e^u - (\alpha e^u)^2}{(\beta + \alpha e^u)^2} = \frac{\alpha e^u}{(\beta + \alpha e^u)}\left(1 - \frac{\alpha e^u}{(\beta + \alpha e^u)}\right)
\end{aligned}$$

Note that $\frac{\alpha e^u}{\beta + \alpha e^u} \in [0, 1]$ since $\alpha, \beta \geq 0$ (this holds because a must be negative and b must be positive due to the condition $\mathbb{E}X = 0$). For $y \in [0, 1]$, the polynomial $y(1 - y)$ takes values in $[0, \frac{1}{4}]$. Therefore, by Taylor's theorem with remainder, we conclude

$$\log \mathbb{E}[e^{tX}] \leq \left(\sup_{u \in \mathbb{R}} \frac{f''(u)}{2}\right) u^2 = \frac{1}{8} u^2 = \frac{t^2(b-a)^2}{8}.$$

□

Question 3. *Let $X_1, \ldots, X_n$ be independent with distribution P on a measurable space $(\mathcal{X}, \mathcal{A})$, and let $\hat{P}_n$ be the empirical measure of $X_1, \ldots, X_n$; thus $\hat{P}_n(A) = n^{-1} \sum_{i=1}^n \mathbb{1}_{X_i \in A}$. Show that, for all $\varepsilon > 0$ and $A \in \mathcal{A}$, we have*

$$\mathbb{P}\left(\left|\hat{P}_n(A) - P(A)\right| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

Proof. Define a new distribution $Y = \mathbb{1}_{X \notin A}$. Its distribution function is given by

$$F_Y(y) = \begin{cases} 0 & y < 0; \\ P(A) & y \in [0, 1);, \\ 1 & y \geq 1. \end{cases} \quad F_Y(0) = P(A).$$

The empirical distribution function of $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} Y$ is given by

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq y},$$

and thus for $y = 0$ we have

$$\hat{F}_n(0) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq 0} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \in A} = \hat{P}_n(A).$$

The result now follows from the DKW inequality. TODO: make more precise. □