Topics in Statistical Theory — Example Sheet 1

Lucas Riedstra

...

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

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

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

Solution. We compute the density function of $U_{(j)}$ as follows: let $x \in (0,1)$, then we know that

$$f_{(j)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{(j)}(x) = \lim_{h \to 0} \frac{F_{(j)}(x+h) - F_{(j)}(x)}{h} = \lim_{h \to 0} \frac{\mathbb{P}(x < U_{(j)} \le x+h)}{h}.$$

The probability $\mathbb{P}(x < U_{(j)} \le x+h)$ is the probability that exactly $j-1$ of the U_i are less than x, and that at least one of the U_i is in $(x, x+h]$.

The probability that two or more of the U_i lie in $(x, x+h]$ is $O(h^2)$ and therefore negligible, so we must compute the probability that exactly $j-1$ of the U_i are smaller than x, one of the U_i is in $(x, x+h]$, and the other U_i are greater than $x+h$. This is easily seen to be

$$\begin{aligned}
&\binom{n}{j-1} \mathbb{P}(U \le x)^{j-1} \cdot (n-j+1) \mathbb{P}(x < U \le x+h) \cdot \mathbb{P}(U > x+h)^{n-j} \\
&= \frac{n!}{(j-1)!(n-j+1)!} (n-j+1) x^{j-1} h (1-x-h)^{n-j} \\
&= \frac{n!}{(j-1)!(n-j)!} x^{j-1} (1-x-h)^{n-j} h.
\end{aligned}$$

Therefore, we easily compute

$$f_{(j)}(x) = \lim_{h \to 0} \frac{\frac{n!}{(j-1)!(n-j)!} x^{j-1} (1-x-h)^{n-j} h}{h} = \frac{n!}{(j-1)!(n-j)!} x^{j-1} (1-x)^{n-j}.$$

This is also the density function of a $\operatorname{Beta}(j, n-j+1)$ distribution.

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

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

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 $\hat{P}_n(A) = n^{-1} \sum_{i=1}^n \mathbb{1}_{U_i \in A}$. 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

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

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 $y \in [0,1)$ we have

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

By the DKW inequality we find

$$\mathbb{P}\Big(\Big|\hat{P}_n(A) - P(A)\Big| > \varepsilon\Big) = \mathbb{P}\Big(\sup_{y\in\mathbb{R}} \Big|\hat{F}_n(y) - F(y)\Big| > \varepsilon\Big) \le 2e^{-2n\varepsilon^2}.$$

$\square$

Question 4. *Let* $X \sim \mathrm{Bin}(n,p)$. *Compare the Hoeffding, Bennett, and Bernstein upper bounds on* $\mathbb{P}(X/n \ge \frac{1}{2})$ *as* $p \to 0$.

Solution. Note that X/n is the average of n i.i.d. random variables $Y_i \sim \mathrm{Bern}(p)$, where $Y_i \in [0,1]$ for all i.

We start with Hoeffding's inequality. In this case, we have

$$\mathbb{P}(X/n - p \ge x) \le \exp\Big(-\frac{2n^2(1/2)^2}{n}\Big) = \exp\Big(-\frac{n}{2}\Big),$$

and this bound also holds as $p \to 0$.

We continue with Bennett's inequality. We consider the random variables $Y_i - p$, which are bounded from above by $b = 1-p$. Now Bennett's inequality tells us, with $\sigma_p^2 = \mathrm{Var}(Y_i - p) = p(1-p)$ that

$$\mathbb{P}(X/n \ge x) \le \exp\Big(-\frac{n}{(1-p)^2}h\Big(\frac{1-p}{2p(1-p)}\Big)\Big) = \exp\Big(-\frac{n}{(1-p)^2}h\Big(\frac{1}{2p}\Big)\Big).$$

We compute

$$h\Big(\frac{1}{2p}\Big) = (1+\frac{1}{2p})\log\Big(1+\frac{1}{2p}\Big) - \frac{1}{2p} = \log\Big(1+\frac{1}{2p}\Big) + \frac{1}{2p}\Big(\log\Big(1+\frac{1}{2p}\Big) - 1\Big) \overset{p\downarrow 0}{\to} +\infty.$$

Since $\frac{n}{(1-p)^2}$ is clearly bounded by n, we conclude that

$$\mathbb{P}(X/n \ge x) \to e^{-\infty} = 0.$$

We finish with Bernstein's inequality. We have for $q \ge 3$ and $p \le \frac{1}{2}$ that

$$\begin{aligned}\mathbb{E}[(Y_i - p)_+^q] &= p(1-p)^q = \sigma_p^2(1-p)^{q-1} = (q!\sigma_p^2(1-p)^{q-2}/2)\cdot(2(1-p)/q!)\\ &\le q!\sigma_p^2((1-p)/3)^{q-2}/2.\end{aligned}$$

Now Bernstein's inequality tells us that

$$\mathbb{P}(X/n \ge \frac{1}{2} \le \exp\Big(-\frac{n(1/2)^2}{2(\sigma_p^2 + (1-p)/6)}\Big) = \exp\Big(-\frac{n}{8\sigma_p^2 + 4(1-p)/3}\Big) \overset{p\to 0}{\to} \exp\Big(-\frac{3n}{4}\Big).$$

Of course, the true limit is 0 for any n, which is only given by Bennett's inequality. We also see that Hoeffding's inequality gives the worst result.

Question 5. *Derive the following alternative form of Bernstein's inequality: under the same conditions,*

$$\mathbb{P}\Big(\bar{X} \ge \sqrt{\frac{2\sigma^2\log(1/\delta)}{n}} + \frac{c}{n}\log(1/\delta)\Big) \le \delta$$

for every $\delta \in (0,1]$.

Proof. We have

$$\mathbb{P}(\bar{X} \geq x) \leq \exp\left(-\frac{nx^2}{2(\sigma^2 + cx)}\right) =: \delta.$$

Now we just need express x in terms of δ: taking logarithms on both sides we obtain

$$-\frac{nx^2}{2(\sigma^2 + cx)} = \log(\delta) \implies nx^2 = 2(\sigma^2 + cx)\log(1/\delta) \implies nx^2 - 2c\log(1/\delta)x - 2\sigma^2\log(1/\delta) = 0.$$

Using the abc-formula with the fact that $x \geq 0$ yields

$$\begin{aligned} x &= \frac{2c\log(1/\delta) + \sqrt{4c^2\log^2(1/\delta) + 8n\sigma^2\log(1/\delta)}}{2n} \\ &= \frac{c}{n}\log(1/\delta) + \sqrt{\frac{c^2}{n^2}\log^2(1/\delta) + \frac{2\sigma^2}{n}\log(1/\delta)} \\ &= ???? \end{aligned}$$

□

Question 6. *(a) Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} F$ and let $\hat{F}_n$ denote their empirical distribution function. For $t_1 < \dots < t_k$, write down the distribution of*

$$n\Big(\hat{F}_n(t_1), \hat{F}_n(t_2) - \hat{F}_n(t_1), \dots, \hat{F}_n(t_k) - \hat{F}_n(t_{k-1}), 1 - \hat{F}_n(t_k)\Big).$$

(b) Find the asymptotic distribution of $n^{1/2}(\hat{F}_n(t_1) - F(t_1), \dots, \hat{F}_n(t_k) - F(t_k))$.

Solution. (a) Write $n\hat{F}_n(t) = \sum_{i=1}^n \mathbb{1}_{X_i \leq t} = \#\{i \mid X_i \leq t\}$, and analogously, for $t < u$, $n(\hat{F}_n(u) - \hat{F}_n(t)) = \#\{i \mid t < X_i \leq u\}$.

Then, defining $t_0 = -\infty$ and $t_{k+1} = \infty$, we find that

$$\begin{aligned} &\mathbb{P}\Big[n\Big(\hat{F}_n(t_1), \dots, 1 - \hat{F}_n(t_k)\Big)(a_1, \dots, a_{n+1})\Big] \\ &= \mathbb{P}[\text{exactly } a_i \text{ of the } X_i \text{ lie in } (t_{i-1}, t_i] \text{ for } i = 1, \dots, n]. \end{aligned}$$

This probability is 0 unless all a_i are nonnegative integers with $\sum_i a_i = n$. In this case, we can first choose a_1 of the X_i which must lie in $(\infty, t_1]$, then we choose a_2 of the remaining X_i which must lie in $(t_1, t_2]$ and so forth. We find that the above probability equals

$$\binom{n}{a_1}(F(t_1))^{a_1}\binom{n - a_1}{a_2}(F(t_2) - F(t_1))^{a_2} \cdots \binom{a_k + a_{k_1}}{a_k}(F(t_k) - F(t_{k-1}))^{a_k} \cdot (1 - F(t_k))^{a_{k+1}}.$$

(b) By the central limit theorem, the asymptotic distribution is $N(0, \Sigma)$, where Σ is the covariance matrix of $(\hat{F}_n(t_1), \dots, \hat{F}_n(t_k))$. We will compute the entries of Σ.

Choose $t \in \mathbb{R}$ arbitrarily. Then we have

$$\begin{aligned} \operatorname{Var}(\hat{F}_n(t)) = \mathbb{E}[\hat{F}_n^2(t)] - \mathbb{E}[\hat{F}_n(t)]^2 &= \mathbb{E}\left[\left(\frac{1}{n}\sum_i \mathbb{1}_{X_i \leq t}\right)^2\right] - F^2(t) \\ &= \frac{1}{n^2}\mathbb{E}\left[\sum_i \mathbb{1}_{X_i \leq t} + 2\sum_{i<j}\mathbb{1}_{X_i \leq t}\mathbb{1}_{X_j \leq t}\right] - F^2(t) \\ &= \frac{F(t) + (n-1)F^2(t)}{n} - F^2(t) = \frac{F(t)(1 - F(t))}{n}, \end{aligned}$$

so we have computed the diagonal entries $\Sigma_{ii} = \frac{F(t_i)(1-F(t_i))}{n}$.
Now we must compute the covariances: assume $s < t$, then

$$\begin{aligned}
\mathrm{Cov}(\hat{F}_n(s), \hat{F}_n(t)) &= \mathbb{E}[\hat{F}_n(s)\hat{F}_n(t)] - \mathbb{E}[\hat{F}_n(s)]\mathbb{E}[\hat{F}_n(t)] \\
&= \frac{1}{n^2}\sum_{i,j} \mathbb{E}[\mathbb{1}_{X_i \le s}\mathbb{1}_{X_j \le t}] - F(s)F(t) \\
&= \frac{1}{n^2}(nF(s) + n(n-1)F(s)F(t)) - F(s)F(t) \\
&= \frac{F(s) + (n-1)F(s)F(t)}{n} - F(s)F(t) = \frac{F(s) - F(s)F(t)}{n}.
\end{aligned}$$

This gives the diagonal entries $\Sigma_{ij} = \frac{F(t_i) - F(t_i)F(t_j)}{n}$ for $i < j$. In the end, we find

$$\Sigma_{ij} = \frac{1}{n} \cdot \begin{cases} F(t_i)(1 - F(t_i)) & \text{if } i = j, \\ F(t_{\min(i,j)}) - F(t_i)F(t_j) & \text{if } i \neq j. \end{cases}$$