170S Week 1 Discussion Notes

Colin Ni

January 8, 2025

Warm-up

The Beta distribution $\mathrm{Beta}(\alpha, \beta)$ for parameters $\alpha > 0$ and $\beta > 0$ is defined by the pdf

$$f_{\mathrm{Beta}(\alpha,\beta)}(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad \text{for } 0 \le x \le 1.$$

Here the $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ term is a normalizing constant so that $\int_0^1 f_{\mathrm{Beta}(\alpha,\beta)}\,\mathrm{d}x = 1$, and $\Gamma(x)$ is the Gamma function which we recall generalizes the factorial function via $\Gamma(n+1) = n!$ and satisfies $\Gamma(x+1) = x\Gamma(x)$. Show that the mean of $\mathrm{Beta}(\alpha, \beta)$ is $\frac{\alpha}{\alpha+\beta}$.

Solution. At once

$$\begin{aligned}
\mathbb{E}(\mathrm{Beta}(\alpha,\beta)) &= \int_0^1 x f_{\mathrm{Beta}(\alpha,\beta)}(x)\,\mathrm{d}x && \text{(definition of mean)} \\
&= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha}(1-x)^{\beta-1}\,\mathrm{d}x && \text{(definition of } \mathrm{Beta}(\alpha,\beta)\text{)} \\
&= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha+1+\beta)\Gamma(\alpha)} \int_0^1 \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{\alpha}(1-x)^{\beta-1}\,\mathrm{d}x \\
& && \text{(massage the integrand to become } f_{\mathrm{Beta}(\alpha+1,\beta)}\text{)} \\
&= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha+1+\beta)\Gamma(\alpha)} \\
& && \text{(the integral is 1 because we made the integrand } f_{\mathrm{Beta}(\alpha+1,\beta)}\text{)} \\
&= \frac{\alpha\Gamma(\alpha)\Gamma(\alpha+\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)} \\
& && \text{(using the reminder } \Gamma(x+1) = x\Gamma(x)\text{)} \\
&= \frac{\alpha}{\alpha+\beta}. && \square
\end{aligned}$$

Intro to Order Statistics

Given some sampled values, the order statistics are the values put in order.

Definition. Let $X_1, \dots, X_n$ be independent samples from a random variable X. The rth order statistic $X_{(r)}$, where $1 \le r \le n$, of these samples is the rth smallest value, or in other words $X_{(1)} \le \cdots \le X_{(n)}$.

Example. If $X_1 = 0.80$, $X_2 = 0.31$, and $X_3 = 0.50$ were sampled from the uniform distrubtion $\mathrm{Unif}(0,1)$ on $[0,1]$, then $X_{(1)} = 0.31$, $X_{(2)} = 0.50$, and $X_{(3)} = 0.80$.

The pdf of $X_{(r)}$ has an extremely nice description.

Key Fact. The pdf of $X_{(r)}$ is

$$f_{X_{(r)}}(x) = r\binom{n}{r} F_X(x)^{r-1}(1 - F_X(x))^{n-r} f_X(x).$$

The slogan to keep in mind is as follows:

Slogan. The pdf of $X_{(r)}$ is a nice polynomial of the cdf and pdf X.

Let us prove this key fact, because the lectures will likely skip this proof.

Proof. The cdf of $X_{(r)}$ is

$$\begin{aligned}
F_{X_{(r)}}(x) &= \mathbb{P}(X_{(r)} \le x) && \text{(by definition)}\\
&= \mathbb{P}(\text{at least } r \text{ of the } n \text{ samples are} \le x) && \text{(think carefully)}\\
&= \sum_{j=r}^{n} \binom{n}{j} F_X(x)^j (1 - F_X(x))^{n-j}.\\
& \text{(these are Bernoulli trials with success probability } \mathbb{P}(X_i \le x) = F_X(x))
\end{aligned}$$

Therefore, taking the derivative gives the pdf

$$\begin{aligned}
f_{X_{(r)}}(x) &= \frac{\mathrm{d}F_{X_{(r)}}}{\mathrm{d}x}(x) && \text{(an important fact from probability)}\\
&= \sum_{j=r}^{n} \binom{n}{j} \frac{\mathrm{d}}{\mathrm{d}x}\left[F_X(x)^j (1 - F_X(x))^{n-j}\right]\\
& && \text{(using our determination of the cdf)}\\
&= \sum_{j=r}^{n} \binom{n}{j} j F_X(x)^{j-1} f_X(x) (1 - F_X(x))^{n-j}\\
&\quad + \sum_{j=r}^{n} \binom{n}{j} F_X(x)^j (n-j)(1 - F_X(x))^{n-j-1}(-f_X(x)).\\
& \text{(product rule and chain rule, recalling again that } \tfrac{\mathrm{d}F_X}{\mathrm{d}x}(x) = f_X(x))
\end{aligned}$$

To simplify this expression, observe that since $\binom{n}{j} = \frac{n!}{j!(n-j)!}$, we have the identities

$$\begin{aligned}
\binom{n}{j} j &= \frac{n!}{(j-1)!(n-j)!} && \text{for } 0 < j \le n\\
\binom{n}{j}(n-j) &= \frac{n!}{j!(n-j-1)!} && \text{for } 0 \le j < n,
\end{aligned}$$

where we need to be careful with the bounds for j so that the expressions $(j-1)!$ and $(n-j-1)!$ make sense. For the first sum, we can plug the first identity without worry because $1 \leq r \leq n$, and for the second sum, the term with $j = n$ is zero anyway. Thus we get

$$\sum_{j=r}^{n} \frac{n!}{(j-1)!(n-j)!} F_X(x)^{j-1}(1-F_X(x))^{n-j} f_X(x) \\ - \sum_{j=r}^{n-1} \frac{n!}{j!(n-j-1)!} F_X(x)^{j}(1-F_X(x))^{n-j-1} f_X(x).$$

Replacing j with $j+1$ in the second sum and changing the bounds to $j = r+1$ to n shows that all terms cancel except for the $j = r$ term in the first sum, which is

$$\frac{n!}{(r-1)!(n-r)!} F_X(x)^{r-1}(1-F_X(x))^{n-r} f_X(x),$$

as desired. $\square$

Fun puzzle / interview problem

Puzzle. Sample 10 values $X_1, \dots, X_{10}$ uniformly on $[0,1]$. What is the expected difference between the 6th and 5th smallest values?

Solution. We want $\mathbb{E}(X_{(6)} - X_{(5)})$. Here our $n = 10$ samples are generated from $\mathrm{Unif}(0,1)$ which has $f_{\mathrm{Unif}(0,1)}(x) = 1$ and $F_{\mathrm{Unif}(0,1)}(x) = x$. By our Key Fact, the rth order statistic has pdf

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} x^{r-1}(1-x)^{n-r}.$$

This is precisely the pdf of $\mathrm{Beta}(r, n-r+1)$! Using the warm-up, the mean of this is $\frac{r}{n+1}$, so

$$\mathbb{E}(X_{(6)} - X_{(5)}) = \mathbb{E}(X_{(6)}) - \mathbb{E}(X_{(5)}) = \frac{6}{11} - \frac{5}{11} = \frac{1}{11}. \qquad \square$$

Here is the important takeaway.

Slogan. The order statistics for Unif are Beta.

Challenge. Come up with a more intuitive solution to this puzzle, using the following idea: uniformly sampling 10 points on $[0,1]$ is the same thing as uniformly sampling 11 points on a circle.