MAT235: Discussion 4

1 Order statistics

Let $X_1, \ldots, X_n$ be independent independent identically distributed variables with a common density function $f$. Let $X_1 \leq \cdots \leq X_n$ denote the order statistics. Then the density of $X_n$ can be obtained as:

$$P(X_n \leq x) = P(\max(X_1, \cdots, X_n) \leq x) = P(X_1 \leq x, \cdots, X_n \leq x) = P(X \leq x)^n = F(x)^n$$

$$\Rightarrow f_{X_n}(x) = \frac{dF(x)^n}{dx} = nF(x)^{n-1}f(x).$$

The density of $X_1$ could also be obtained as:

$$P(X_1 \geq x) = P(\min(X_1, \cdots, X_n) \geq x) = P(X_1 \geq x, \cdots, X_n \geq x) = P(X \geq x)^n = (1-F(x))^n$$

$$\Rightarrow f_{X_1}(x) = \frac{d[1-(1-F(x))^n]}{dx} = n(1-F(x))^{n-1}f(x).$$

For general $X_{(k)}$, the density is:

$$f_{X_{(k)}}(x)dx = P(X_{(k)} \in [x, x + dx)) = P(\text{one of } X \text{'s } \in [x, x + dx), \ k-1 \text{ of } X \text{'s } < x$$

$$= \binom{n}{1} P(X \in [x, x + dx)) \binom{n-1}{k-1} P(X < x)^{k-1}P(X \geq x)^{n-k}$$

$$= nf(x)dx \cdot \binom{n-1}{k-1} F(x)^{k-1}(1-F(x))^{n-k}$$

$$\Rightarrow f_{X_{(k)}}(x) = nf(x) \binom{n-1}{k-1} F(x)^{k-1}(1-F(x))^{n-k}.$$

For $1 \leq i < j \leq n$, the joint density of $X_{(i)}, X_{(j)}$ can be obtained similarly:

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} f(u)f(v)F(u)^{i-1}[F(v) - F(u)]^{j-i-1}[1 - F(v)]^{n-j},$$

if $u < v$. Generalize this result to the joint density of $X_{(i_1)}, \cdots, X_{(i_k)}$ for $1 \leq i_1 < \cdots < i_k \leq n$.

**Example:** Let $X_1, \cdots, X_n$ be independent $U(0,1)$. The density of $X_{(k)}$ is

$$\frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k},$$

which is a beta distribution $Beta(k, n-k+1)$. What is the density of $nX_{(n)}$ and what is the limit of this density function as $n \to \infty$?

**Example:** Let $X_1, X_2, X_3$ be independent $U(0,1)$. What is the probability that the rods of length $X_1, X_2, X_3$ may be used to make a triangle.
A triangle may be formed if $X_{(3)} < X_{(1)} + X_{(2)}$. And by previous result we know the joint density of $X_{(1)}, X_{(2)}, X_{(3)}$ is $f_{X_{(1)}, X_{(2)}, X_{(3)}}(u, v, w) = 6$ if $0 \leq u \leq v \leq w \leq 1$. Therefore:

$$
P(\text{triangle}) = 1 - P(X_{(3)} \geq X_{(1)} + X_{(2)})
= 1 - \int_0^1 \left[ \int_0^{u/2} \int_0^v 6dw dv + \int_{u/2}^u \int_0^{u-v} 6dw dv \right] du
= 1 - 6 \int_0^1 \left[ \int_0^{u/2} vdv + \int_{u/2}^u (u-v)dv \right] du
= 1 - 6 \int_0^1 \left[ \frac{u^2}{8} + \frac{u^2}{8} \right] du
= 1 - 6 \frac{1}{12} = \frac{1}{2}.
$$

## 2 Poisson process

**Definition:** A Poisson process with rate $r$ is a process $N_t = \{N(t) : t \geq 0\}$ taking values in $0, 1, 2, \ldots$ such that:

1. $N(0) = 0$;
2. $N(t)$ has independent increments, i.e. $N(t_i) - N(t_{i-1})$ are independent for $i = 1, 2, 3, \ldots$;
3. $N(t) - N(s) \sim \text{Poi}(r(t-s))$.

**Theorem:** Let $T_i$ denote the waiting time of the $i$-th arrival i.e. $T_0 = 0$, $T_i = \inf\{t : N(t) = i\}$. The interarrival times are denoted by $X_1, X_2, \ldots$, i.e. $X_i = T_i - T_{i-1}$. Then $X_1, X_2, \ldots$ are independent, each having the exponential distribution with parameter $r$ if $N(t)$ is a Poisson process with rate $r$.

First consider $X_1$:

$$
P(X_1 > t) = P(N(t) = 0) = e^{-rt}.
$$

Now conditioning on $X_1$,

$$
P(X_2 > t | X_1 = t_1) = P(\text{no arrivals in} (t_1, t_1 + t] | X_1 = t_1).
$$

Where $\{X_1 = t_1\} = \{N(t_1) - N(0) = 1\}$ and $\{\text{no arrivals in} (t_1, t]\} = \{N(t) - N(t_1) = 0\}$ are independent by the definition.

$$
P(X_2 > t | X_1 = t_1) = P(\text{no arrivals in} (t_1, t_1 + t]) = e^{-rt}.
$$

Similarly,

$$
P(X_{n+1} > t | X_1 = t_1, \ldots, X_n = t_n) = P(\text{no arrivals in} (T, T + t]) = e^{-rt},
$$

where $T = t_1 + \cdots + t_n$.

**Theorem:** Let $T_i, X_i$ be the same as in above theorem. Then $N(t) = \max\{n : T_n \leq t\}$ is a Poisson random variable with parameter $rt$, if $X_1, X_2, \ldots$ are independent identically distributed $\text{Exp}(r)$.

Notice that $T_n = \sum_{i=1}^n X_i$, and $\{N(t) \geq n\} = \{T_n \leq t\}$,

$$
P(N(t) = n) = P(T_n \leq t < T_{n+1}) = P(T_n \leq t) - P(T_{n+1} \leq t).
$$
And $T_n \sim \text{Gamma}(n, r)$,

$$P(N(t) = n) = \int_0^t \frac{r^n}{n!} x^{n-1} e^{-rx} \, dx - \int_0^t \frac{r^{n+1}}{(n+1)!} x^n e^{-rx} \, dx$$

$$= \frac{(rt)^n}{n!} e^{-rt}.$$

by integration by parts and induction.

**Example:** Simulate $Poi(\lambda)$ using $U_1, U_2, \cdots \sim \text{iid} U(0, 1)$.

Using the fact $T_n = \sum_{i=1}^n X_i$ and $N(t) \sim Poi(\lambda t)$,

$$P(N(t) \geq n) = P(T_n \leq t) = P(\sum_{i=1}^n X_i \leq t),$$

$$\{\prod_{i=1}^n U_i \leq e^{-\lambda t}\} \Leftrightarrow \{-\sum_{i=1}^n \log U_i \geq \lambda t\}.$$