Lecture 10.

Order Statistics.

The order statistics of a set of r.v's $X_1, X_2, \dots, X_n$ are the same r.v's arranged in increasing order.

$X_1, \dots, X_n \sim$ same cdf $F(x)$, $f(x)$

Denote by

$X_{(1)} =$ smallest of $X_1, \dots, X_n$

$X_{(1)} = \min\{X_1, \dots, X_n\}$ $\qquad X_1 \neq X_{(1)}$

$X_{(2)} = 2^{nd}$ smallest

$\vdots$

$X_{(n)} =$ largest of $X_1, \dots, X_n$

$X_{(n)} = \max\{X_1, \dots, X_n\}$

$X_1, \dots, X_n$ are independent $\not\Rightarrow X_{(1)}, \dots, X_{(n)}$ are indep

because $X_{(1)} \le \dots \le X_{(n)}$

Distribution of the Largest order statistics $X_{(n)}$.

Suppose $X_1, X_2, \dots, X_n$ are i.i.d. r.v's with common distribution function $F_X(x)$ and density function $f_X(x)$.

$$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(X_1 \le x, \dots, X_n \le x)$$

$$= P(X_1 \le x) \cdots P(X_n \le x)$$

$$= [F(x)]^n$$

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

<u>Ex.</u> $X_1, X_2, \dots, X_n \sim$ iid Unif$(0,1)$. Find $f_{X_{(n)}}(x)$.

$$F_{X_i}(x) = x, \quad x \in (0,1)$$

$$f_{X_i}(x) = \begin{cases} 1, & x \in (0,1) \\ 0, & \text{ow} \end{cases}$$

$$f_{X_{(n)}}(x) = n\left[F_{X_i}\right]^{n-1} f_{X_i} = \begin{cases} n\,x^{n-1}, & x \in (0,1) \\ 0, & \text{ow} \end{cases}$$

Distribution of the Smallest orderstatistic $X_{(1)}$.

Suppose $X_1, X_2, \dots, X_n$ are i.i.d. r.v's with common distribution function $F_X(x)$ and density function $f_X(x)$.

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \le x) = 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x, \dots, X_n > x) \\ &= 1 - P(X_1 > x) \cdots P(X_n > x) \\ &= 1 - [1 - F(x)]^n \end{aligned}$$

$$f_{X_{(1)}}(x) = \frac{d}{dx} F_{X_{(1)}}(x) = n(1 - F(x))^{n-1} \cdot f(x)$$

<u>Ex</u>. $X_1, X_2, \ldots, X_n \sim iid$ Unif $(0,1)$. Find $f_{X_{(1)}}(x)$.

$$f_{X_i} = 1, x\,(0,1)$$

$$F_{X_i} = x, x \in (0,1)$$

$$f_{X_{(1)}}(x) = \begin{cases} n(1-x)^{n-1}, & x \in (0,1) \\ 0, & ow \end{cases}$$

Distribution of the k^{th} order statistic $X_{(k)}$.

Suppose $X_1, X_2, \ldots, X_n$ are iid r.v's with common $F_X(x)$ and $f_X(x)$.

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!}\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k} f(x)$$

$$P(x \le X_{(k)} \le x+dx)$$

$k-1$ r.v's are below x $\sim F(x)^{k-1}$

$n-k$ r.v's are above $x+dx$ $\sim \left[1-F(x+dx)\right]^{n-k}$

and there is one point inside $(x, x+dx)$

$\sim \underline{P(x \le X \le x+dx)}$

Ex. $X_1, X_2, \dots, X_n \sim$ iid $\text{Unif}(0,1)$. Find $f_{X_{(k)}}(x)$.

$$f_{X_{(k)}}(x) = \begin{cases} \dfrac{n!}{(k-1)!\,(n-k)!}\, x^{k-1}(1-x)^{n-k}, & x \in (0,1) \\ 0, & \text{ow} \end{cases}$$

Generating Functions.

For a sequence of real numbers $\{a_j\} = a_0, a_1, a_2, \dots$, the generating function of $\{a_j\}$ is

$$A(t) = \sum_{j=0}^{\infty} a_j t^j$$

if this converges for $|t| < t_0$ for some $t_0 > 0$.

$$A(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \dots$$

$$A(0) = a_0$$

$$A'(t) = a_1 + 2a_2 t + 3a_3 t^2 + \dots$$

$$A'(0) = a_1$$

$$A''(t) = 2a_2 + 6a_3 t + \dots$$

$$A''(0) = 2a_2 \quad \dots$$

Probability Generating Functions

Suppose X is a r.v. taking the values $0, 1, 2, \ldots$

Let $p_j = P(X=j)$, $j = 0, 1, 2, \ldots$

<u>Def.</u> The **probability generating function** of X is

$$\pi_X(t) = p_0 + p_1 t + p_2 t^2 + \ldots = \sum_{j=0}^{\infty} p_j t^j$$

$G_X(t)$

$P_X(t)$ <u>Note</u>: $|p_j t^j| \le p_j$ if $|t| < 1$

$$\underset{|t|<1}{\sum |p_j t^j| \le \sum p_j = 1}$$

So $\pi_X(t)$ converges absolutely at least for $|t| < 1$.

$$\pi_X(t) = \sum_{j=0}^{\infty} p_j t^j = E(t^X), \quad \pi_X(1) = 1$$

pgf

<u>Ex.</u> $X \sim Bin(n, p)$

$$\pi_X(t) = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} t^j = \sum_{j=0}^{n} \binom{n}{j} (pt)^j q^{n-j}$$

$$= (pt+q)^n, \quad t \in \mathbb{R}$$

$X \sim Geom(p)$

$$\pi_X(t) = \sum_{j=1}^{\infty} p q^{j-1} t^j = \frac{p}{q} \sum_{j=1}^{\infty} (qt)^j$$

$$= \frac{p}{\cancel{q}} \cdot \frac{\cancel{q}t}{1-qt}, \quad |qt| < 1$$

$$= \frac{pt}{1-qt}, \quad |t| < \frac{1}{q}$$

PGF for sums of independent r.v's.

Let X, Y be independent, $Z = X + Y$

$$\pi_Z(t) = E(t^Z) = E(t^{X+Y}) = E(t^X t^Y) = E(t^X)E(t^Y)$$
$$= \pi_X(t)\,\pi_Y(t)$$

Ex. Let $Y = \mathrm{Bin}(n, p)$

$Y = X_1 + \dots + X_n$, where $X_i \sim$ iid Bernoulli(p)

$$X_i = \begin{cases} 1 & \text{with } p \\ 0 & \text{with } 1-p \end{cases}$$

$$\pi_{X_i}(t) = t^0(1-p) + t^1 p$$
$$= tp + q$$

$$\pi_Y(t) = \pi_{X_1}(t) \cdots \pi_{X_n}(t)$$
$$= (tp + q)^n$$

Theorem. Let X be a discrete r.v, whose possible values are the non-negative integers. Assume $\Pi_X(t_0) < \infty$ for some $t_0 > 0$. Then

$$\Pi_X(0) = P(X=0),$$
$$\Pi_X'(0) = P(X=1),$$
$$\Pi''_X(0) = 2P(X=2),$$
etc.

In general,

$$\Pi_X^{(k)}(0) = k!\,P(X=k),$$

where $\Pi_X^{(k)}$ is the k^{th} derivative of Π_X w.r. to t.

Proof:

$$\Pi_X(t) = p_0 + p_1 t + p_2 t^2 + \ldots$$
$$\Pi_X(0) = p_0 = P(X=0)$$
$$\Pi'(t) = p_1 + 2p_2 t + \ldots$$
$$\Pi'(0) = p_1 = P(X=1)$$
$$\Pi''(t) = 2p_2 + 6p_3 t + \ldots$$
$$\Pi''(0) = 2p_2 = 2P(X=2)$$
$$\vdots$$

Ex. $X \sim \text{poisson}(\lambda)$.

$$\pi_X(t) = \sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!} t^j = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda t)^j}{j!}$$

$$= e^{-\lambda} e^{\lambda t} = \underline{e^{\lambda(t-1)}}$$

$$\pi_X(0) = e^{-\lambda} = \frac{e^{-\lambda}\lambda^0}{0!} = P(X=0)$$

$$\pi_X'(t) = \lambda e^{\lambda(t-1)}$$

$$\pi_X'(0) = \lambda e^{-\lambda} = \frac{\lambda^1 e^{-\lambda}}{1!} = P(X=1)$$

$$\pi_X''(t) = \lambda^2 e^{\lambda(t-1)}$$

$$\pi_X''(0) = \lambda^2 e^{-\lambda} = 2 \cdot \frac{\lambda^2 e^{-\lambda}}{2!} = 2P(X=2)$$

$$\vdots$$

$$\pi_X^{(k)}(t) = \lambda^k e^{\lambda(t-1)}$$

$$\pi_X^{(k)}(0) = \lambda^k e^{-\lambda} = k! \cdot \frac{\lambda^k e^{-\lambda}}{k!}$$

$$= k!\, P(X=k)$$

Theorem. Let X be a discrete r.v., whose possible values are the non-negative integers. If $\Pi_X(t) < \infty$ for $|t| < t_0$ for some $t_0 > 1$. Then

$$\Pi_X'(1) = E(X)$$

$$\Pi_X''(1) = E[X(X-1)]$$

etc.

In general,

$$\Pi_X^{(k)}(1) = E[X(X-1)(X-2)\ldots(X-k+1)].$$

$E(X(X-1)\ldots(X-k+1)) =$ k^{th} factorial moment

Proof:

$$\Pi_X(t) = p_0 + p_1 t + p_2 t^2 + p_3 t^3 + \ldots$$

$$\Pi_X'(t) = p_1 + 2p_2 t + 3p_3 t^2 + \ldots$$

$$\Pi_X(1) = p_1 + 2p_2 + 3p_3 + 4p_4 + \ldots$$

$$= 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + \ldots$$

$$= E(X)$$

$$\Pi''(t) = 2p_2 + 6p_3 t + 12 p_4 t^2 + \ldots$$

$$\Pi''(1) = 2p_2 + 6p_3 + 12p_4 + \ldots$$

$$= (0^2-0)p_0 + (1^2-1)p_1 + (2^2-2)p_2$$

$$+ (3^2-3)p_3 + (4^2-4)p_4 + \ldots$$

$$= E(X^2-X) = E[X(X-1)]$$

Ex. $X \sim Bin(n,p)$. Find $E(X)$ and $Var(X)$ using the pgf of X, $\Pi_X(t) = (pt+q)^n$.

$$E(X) = \Pi_X'(1) = n(p+q)^{n-1}p = np$$

$$\Pi_X'(t) = n(pt+q)^{n-1}p$$

$$Var(X) = E(X^2) - E(X)^2 = E[X(X-1)] + E(X) - E(X)^2$$

$$= \Pi_X''(1) + \Pi_X'(1) - \Pi_X'(1)^2$$

$$\Pi_X''(t) = n(n-1)(pt+q)^{n-2}p^2\Big|_{t=1} = n(n-1)p^2$$

$$Var(X) = n(n-1)p^2 + np - n^2p^2$$

$$= -np^2 + np = np(1-p) = npq$$

Uniqueness Theorem for PGF. Suppose X, Y have pgf's π_X and π_Y respectively. Then

$$\pi_X(t) = \pi_Y(t) \quad \text{iff} \quad P(X=k) = P(Y=k) \text{ for } k = 0,1,2,\ldots$$

Ex: $\pi_X(t) = 0.2 + 0.3t^3 + 0.5t^7$

Y:	0	3	7
	0.2	0.3	0.5

$\pi_Y(t) = \pi_X(t) \Rightarrow X \overset{d}{=} Y$ (X and Y have the same dist'n)

Moment Generating Functions.

The **moment generating function** (mgf) of a r.v. X is

$$m_X(t) = E(e^{tX})$$

$m_X(t)$ exists if $m_X(t) < \infty$ for $|t| < t_0$, $t_0 > 0$

X - discrete

$$m_X(t) = \sum_x e^{tx} p_X(x)$$

Note: $m_X(t) = \pi_X(e^t)$

X - continuous

$$m_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$$

Ex. $X \sim \mathcal{E}xp(\lambda)$.

$$m_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx$$

$$= \lambda \int_0^\infty e^{-x(\lambda-t)} dx = \frac{-\lambda}{\lambda-t} e^{-x(\lambda-t)} \Big|_0^\infty = \frac{\lambda}{\lambda-t}$$

$$t < \lambda$$

$X \sim \text{Unif}(0,1)$.

$$m_X(t) = \int_0^1 e^{tx} dx = \frac{1}{t} e^{tx} \Big|_0^1 = \frac{1}{t}[e^t - 1]$$

$X \sim \text{Poisson}(\lambda)$

$$m_X(t) = \sum_{x=0}^\infty e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^\infty \frac{(e^t \lambda)^x}{x!}$$

$$= e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t - 1)}$$

<u>Theorem</u>. Let X be any r.v. If $m_X(t) < \infty$ for $|t| < t_0$, $t_0 > 0$, then

$$m_X(0) = 1$$
$$m_X'(0) = E(X)$$
$$m_X''(0) = E(X^2)$$
etc.

In general,

$$m_X^{(k)}(0) = E(X^k).$$

<u>Proof</u>: Discrete case:

$$m_X(t) = p_0 + e^{t} p_1 + e^{2t} p_2 + \dots$$

$$m_X(0) = p_0 + p_1 + p_2 + \dots = 1$$

$$m_X'(t) = e^{t} p_1 + 2e^{2t} p_2 + 3e^{3t} p_3 + \dots,$$

$$m_X'(0) = p_1 + 2p_2 + 3p_3 + \dots = E(X)$$

$$m_X''(t) = e^{t} p_1 + 4e^{2t} p_2 + 9e^{3t} p_3 + \dots$$

$$m_X''(0) = p_1 + 4p_2 + 9p_3 + \dots$$

$$= \sum_{x} x^2 p_x = E(X^2)$$

∎

Ex. $X \sim Exp(\lambda)$. Find $E(X)$ and $Var(X)$ using its mgf.

$$m_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx$$

$$= \int_0^\infty \lambda e^{-x(\lambda - t)} dx = \frac{\lambda}{\lambda - t}, \quad t < \lambda$$

$$E(X) = m'_X(0) = \frac{\lambda}{(\lambda - t)^2}\Big|_{t=0} = \frac{1}{\lambda}$$

$$E(X^2) = m''_X(0) = \frac{2\lambda}{(\lambda - t)^3}\Big|_{t=0} = \frac{2}{\lambda^2}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Ex. $X \sim N(0,1)$. Find $E(X)$ and $Var(X)$.

$$m_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx - \frac{x^2}{2}} dx$$

$$= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t^2 - 2tx + x^2)} dx$$

$$= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} dx$$

$x - t = u$

$$= e^{t^2/2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} dx}_{=1} = e^{t^2/2}$$

$m_X(t) = e^{t^2/2}$

$E(X) = m_X'(0) = t e^{t^2/2}\Big|_{t=0} = 0$

$E(X^2) = m_X''(0) = e^{t^2/2} + t^2 e^{t^2/2}\Big|_{t=0} = 1$

$Var(X) = 1 - 0^2 = 1$

Ex. $X \sim Bin(n,p)$. Find $E(X)$ and $Var(X)$.

$m_X(t) = \sum_x e^{tx}\binom{n}{x} p^x (1-p)^{n-x} = \sum_x \binom{n}{x}(e^t p)^x (1-p)^{n-x} = [e^t p + (1-p)]^n$

$E(X) = m_X'(0) = n[e^t p + (1-p)]^{n-1} p e^t \Big|_{t=0} = np$

$E(X^2) = m_X''(0) = n(n-1)[e^t p + (1-p)]^{n-2} p^2 e^{2t}$
$\quad + n[e^t p + (1-p)]^{n-1} p e^t \Big|_{t=0}$
$\quad = n(n-1)p^2 + np \Rightarrow Var(X) = n(n-1)p^2 + np - n^2p^2$
$\quad = npq$

Properties of MGF:

- $m_X(0) = 1$
- If $Y = a + bX$, then
  $m_Y(t) = E(e^{tY}) = E(e^{at+btX}) = e^{at} E(e^{btX})$
  $= e^{at} m_X(bt)$
- If X, Y are independent, $Z = X + Y$
  $m_Z(t) = E(e^{tX+tY}) = E(e^{tX})E(e^{tY}) = m_X(t) m_Y(t)$

Uniqueness Theorem: If $m_X(t) = m_Y(t)$ for all values of t, then X and Y have the same probability distribution.

Ex. $Y \sim \text{Geom}(p)$. Find the mgf of Y. 10.17

$$m_Y(t) = \sum_{y=1}^{\infty} e^{ty} p q^{y-1} = pe^t \sum_{y=1}^{\infty} e^{t(y-1)} q^{y-1}$$

$$= pe^t \sum_{y=0}^{\infty} (e^t q)^y = pe^t \cdot \frac{1}{1-e^t q}, \text{ if}$$

$$e^t q < 1 \Rightarrow t < \ln \frac{1}{q}$$

$$m_Y(t) = \frac{pe^t}{1-qe^t}, \quad t < -\ln q$$

Let a r.v. X have mgf $m_X(t) = \dfrac{0.3e^t}{1-0.7e^t}$.

What is the dist'n of X?

By uniqueness thm,

$X \sim \text{Geom}(0.3)$

10/18

<u>Ex</u>. Let $X \sim N(\mu, \sigma^2)$. Find mgf of X.

Suppose $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim (\mu_2, \sigma_2^2)$ are independent. Find the distribution of $Y = X_1 + X_2$.