Chapter 4. Multiple Random Variables

4.1 Joint & marginal distributions

<u>Defn</u>: An n-dim'l random vector is a func. from a sample space S into $\mathbb{R}^n$, n-dim'l Euclidean space.

(e.g.) (Tossing two dice)

X = sum of the two dice, Y = | diff. of the two dice |

(X, Y) is a bivariate random vector.

<u>Defn</u>: Let (X,Y) be a discrete bivariate random vector. Then

$f(x,y) = P(X=x, Y=y)$ is called the joint pmf of (X,Y).

- $A \subset \mathbb{R}^2$, $P((X,Y) \in A) = \sum_{(x,y)\in A} f(x,y)$
- $E\, g(X,Y) = \sum_{(x,y)\in \mathbb{R}^2} g(x,y) f(x,y)$.

Th^{m}: (X, Y), a bivariate discrete random vector with joint pmf $f(x,y)$.

Then
$$\left.\begin{aligned} f_X(x) &= P(X=x) = \sum_{y \in \mathbb{R}} f(x,y), \\ f_Y(x) &= P(Y=y) = \sum_{x \in \mathbb{R}} f(x,y). \end{aligned}\right\} \text{ marginal pmf's}$$

(e.g.) (same marginals, different joint pmf) See Example 4.1.9. in p. 144.

- The joint pmf tells us additional information about the dist^{n} of (X, Y) that can not be found in the marginals.

Defn: A function $f(x,y)$ from $\mathbb{R}^2$ to $\mathbb{R}_+$ is called the joint pdf of the continuous bivariate random vector (X,Y) if, for every $A \subset \mathbb{R}^2$,

$$P((X,Y) \in A) = \iint_A f(x,y)\, dx\, dy.$$

- $E\, g(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy$
- $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$ (the marginals)

(e.g.) $f(x,y) = e^{-y}\, I(0 < x < y < \infty)$

$$P(X+Y \geq 1) = 1 - P(X+Y < 1) = \cdots$$

- $F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\, dv\, du$ (the joint cdf)
- $\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y).$

4.2 Conditional dist^n's and independence.

Dfn: (X, Y), a discrete [conti.] bivariate r. vec.

$$f(x|y) \left[= P(X=x \mid Y=y)\right] = f(x,y)/f_Y(y)$$

(the conditional pmf of X given that $Y=y$).

- $f(x|y)$ is also a pmf.
- $E(g(X) \mid y) = \sum_x g(x) f(x|y)$ or $\int_{-\infty}^{\infty} g(x) f(x|y)\, dx$.

Dfn: $f(x,y) = f_X(x) f_Y(y)$ for all $(x,y) \in \mathbb{R}^2$.

Then X and Y are called independent random variables.

<u>Lemma</u>: $(X, Y) \sim f(x,y)$.

X, Y are indep. $\Longleftrightarrow$ $\exists\, g(x)$ and $h(y)$ s.t. $f(x,y) \overset{\forall x,y}{=} g(x)h(y)$

<u>Thm</u>: X, Y, indep. r.v.'s

$$E\, g(X)h(Y) = E\, g(X) \cdot E h(Y).$$

$$M_{X+Y}(t) = M_X(t)\, M_Y(t)$$

(e.g.) $X \sim N(\mu, \sigma^2)$, $Y \sim N(\gamma, \tau^2)$, indep.

Then $X+Y \sim N(\mu+\gamma, \sigma^2+\tau^2)$.

4.3 Bivariate transformations

(e.g.) $X \sim$ Poisson(θ), $Y \sim$ Poisson(λ), indep.

Then $X+Y \sim$?

sol. $U = X+Y$, $V = Y$

$$f_{U,V}(u,v) = P(X+Y=u,\ Y=v) = P(X=u-v,\ Y=v) = P(X=u-v)P(Y=v)$$

$$= f_X(u-v)\, f_Y(v) = \frac{e^{-\theta}\,\theta^{u-v}}{(u-v)!} \cdot \frac{e^{-\lambda}\lambda^{v}}{v!} = \frac{1}{u!} \cdot e^{-(\theta+\lambda)} \cdot \binom{u}{v} \lambda^{v} \theta^{u-v}$$

$$f_U(u) = \sum_{v=0}^{u \leftarrow \infty} f_{U,V}(u,v)$$

$$= \frac{e^{-(\theta+\lambda)}}{u!} \cdot \underbrace{\sum_{v=0}^{u} \binom{u}{v} \lambda^{v} \theta^{u-v}}_{(\theta+\lambda)^{u}} = \frac{e^{-(\theta+\lambda)}(\theta+\lambda)^{u}}{u!}, \quad u = 0, 1, 2, \cdots$$

: Poisson$(\theta+\lambda)$ //

- (X,Y), conti. with pdf $f_{X,Y}(x,y)$.

$U = g_1(X,Y)$, $V = g_2(X,Y)$.

$\Longrightarrow$ $X = h_1(U,V)$, $Y = h_2(U,V)$

$$f_{U,V}(u,v) = f_{X,Y}\big(h_1(u,v),\ h_2(u,v)\big) \cdot |J|$$

where $J = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$.

(e.g.) $X, Y \overset{iid}{\sim} N(0,1)$. Then $X/Y \sim$?

sol. $U = X/Y$, $V = |Y|$.

(See Example 4.3.6 in page 162),

4.4 Hierarchical models and mixture dist^{n}'s.

(e.g.) (Binomial – Poisson hierarchy) An insect lays a large number of eggs, each surviving w.p. p. On the average, how many eggs will survive?

$X = \#$ of survivors, $Y = \#$ of eggs laid

$$X \mid Y \sim B(Y, p), \quad Y \sim \text{Poisson}(\lambda)$$

- The advantage of the hierarchical model is that complicated processes may be modeled by a seq. of simple models placed in a hierarchy.
- Dealing with the hierarchy is no more difficult than dealing with conditional and marginal dist^{n}'s.

(e.g.) (cont'd)

$$P(X=x) = \sum_{y=0 \to x}^{\infty} P(X=x, Y=y), \quad 0 \le x \le y$$

$$= \sum_{y=0 \to x}^{\infty} P(X=x \mid Y=y)\, P(Y=y)$$

$$= \sum_{y=0 \to x}^{\infty} \binom{y}{x} p^{x} (1-p)^{y-x} \cdot \frac{e^{-\lambda} \lambda^{y}}{y!}$$

$$= \frac{(\lambda p)^{x} e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{\{(1-p)\lambda\}^{y-x}}{(y-x)!} \qquad (y-x=t)$$

$$= \frac{(\lambda p)^{x} e^{-\lambda}}{x!} \sum_{t=0}^{\infty} \frac{\{(1-p)\lambda\}^{t}}{t!}$$

$$= \frac{(\lambda p)^{x} e^{-\lambda}}{x!} \cdot e^{(1-p)\lambda} = \frac{(\lambda p)^{x} e^{-\lambda p}}{x!} \quad : \text{Poisson}(\lambda p) \,/\!/$$

<u>Th^{m}</u> : $E(X) = E\big(E(X|Y)\big)$

(e.g.) (cont'd)

$$EX = E(E(X|Y)) = E(Yp) = \lambda p. \,//$$

<u>Defn</u> : X is said to have a mixture distn if the distn of X depends on a quantity that also has a distn.

<u>Th^{m}</u> : $\operatorname{Var} X = E \operatorname{Var}(X|Y) + \operatorname{Var} E(X|Y)$.

(e.g.) (cont'd)

$$\operatorname{Var} X = E\{Yp(1-p)\} + \operatorname{Var}\{Yp\} = \lambda p(1-p) + \lambda p^2 = \lambda p. \,//$$

4.5 Covariance and correlation

Defn: $\mathrm{Cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$

$$\rho_{XY} = \mathrm{Cov}(X,Y)/\sigma_X\sigma_Y$$

- $\mathrm{Cov}(X,Y) = EXY - \mu_X\mu_Y$

Thm: If X and Y are indep. r.v.'s, then $\mathrm{Cov}(X,Y)=0$ and $\rho_{XY}=0$.

Thm: $\mathrm{Var}(aX+bY) = a^2\,\mathrm{Var}\,X + b^2\,\mathrm{Var}\,Y + 2ab\,\mathrm{Cov}(X,Y)$.

Thm: $|\rho_{XY}| \le 1$.

Def^n : $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y > 0$, $|\rho| < 1$

$$f(x,y) = \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}\right)^{-1} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right\}$$

: bivariate normal pdf. $(X,Y) \sim N_2\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \cdot & \sigma_Y^2\end{pmatrix}\right)$.

- $(X,Y) \sim N_2\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \cdot & \sigma_Y^2\end{pmatrix}\right)$

Then, $aX + bY \sim N\left(a\mu_X + b\mu_Y,\ a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y\right)$.

$X \mid Y=y \sim N\left(\mu_X + \rho(\sigma_X/\sigma_Y)(y-\mu_Y),\ \sigma_X^2(1-\rho^2)\right)$

4.6 Multivariate distⁿ's

$$\underset{\sim}{X} = (X_1, \cdots, X_n)$$

Defn : Let n and m be positive integers, and let $p_1, \cdots, p_n$ be numbers sat'g $0 \le p_i \le 1$ and $\sum_i p_i = 1$. Then,

$\underset{\sim}{X} = (X_1, \cdots, X_n)$ has a multinomial distⁿ with m trials and cell prob.'s $p_1, \cdots, p_n$ if the joint pmf is

$$f(x_1, \cdots, x_n) = \frac{m!}{x_1!\, x_2! \cdots x_n!} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n}$$

on the set $\{ (x_1, \cdots, x_n)$ s.t. $x_i \geq 0$, $\sum_1^n x_i = m$.

- Multinomial theorem :

$$(p_1 + \cdots + p_n)^m = \sum_{\substack{x_i \geq 0 \\ \Sigma x_i = m}} \frac{m!}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_n^{x_n} .$$

4.7 Inequalities

<u>Lemma</u> : $a > 0$, $b > 0$, $p > 1$, $q > 1$. $\frac{1}{p} + \frac{1}{q} = 1$

Then, $\frac{1}{p} a^p + \frac{1}{q} b^q \geq ab$ with equality iff $a^p = b^q$.

Pf. Fix $b > 0$.

$g(a) = \frac{1}{p} a^p + \frac{1}{q} b^q - ab$

$g'(a) = 0 \Rightarrow b = a^{p-1} \Rightarrow \frac{1}{p} a^p + \frac{1}{q} (a^{p-1})^q - a \cdot a^{p-1} = 0$, i.e. minimum $= 0$. //

$\underline{\text{Th}^{m}}$ (Hölder) $p>1,\ q>1$ s.t. $\frac{1}{p}+\frac{1}{q}=1$. Then,

$$|EXY| \overset{\text{trivial}}{\leq} E|XY| \leq \left(E|X|^p\right)^{1/p}\left(E|Y|^q\right)^{1/q}.$$

$\|XY\|_1 \leq \|X\|_p \cdot \|Y\|_q$

Pf $a=\frac{|X|}{(E|X|^p)^{1/p}}$, $b=\frac{|Y|}{(E|Y|^q)^{1/q}}$.

Applying Lemma, we get

$$\frac{1}{p}\cdot\frac{|X|^p}{E|X|^p}+\frac{1}{q}\cdot\frac{|Y|^q}{E|Y|^q} \geq \frac{|XY|}{(E|X|^p)^{1/p}\cdot(E|Y|^q)^{1/q}}$$

$$\underbrace{\frac{1}{p}\cdot\underbrace{\frac{E|X|^p}{E|X|^p}}_{=1}+\frac{1}{q}\cdot\underbrace{\frac{E|Y|^q}{E|Y|^q}}_{=1}}_{1} \geq \frac{E|XY|}{(E|X|^p)^{1/p}\cdot(E|Y|^q)^{1/q}}$$

//

c.f. $p=q=2$: Cauchy–Schwarz.

- (Liapounov's ineq.) Set $Y \equiv 1$ in the Hölder's ineq. Then,

$$E|X| \le \left(E|X|^p\right)^{1/p}, \quad p > 1.$$

For $1 < r < p$, if we replace $|X|$ by $|X|^r$, then

$$E|X|^r \le \left(E|X|^{rp}\right)^{1/p}.$$

$$\left(E|X|^r\right)^{1/r} \le \left(E|X|^{rp}\right)^{1/rp}.$$

s

$$\therefore \left(E|X|^r\right)^{1/r} \le \left(E|X|^s\right)^{1/s}, \quad 1 < r < s$$

$\|X\|_r \le \|X\|_s$

$\underline{\text{Th}^m}$ (Minkowski's ineq.) For $1 \le p < \infty$,

$$\left(E|X+Y|^p \right)^{1/p} \le \left(E|X|^p \right)^{1/p} + \left(E|Y|^p \right)^{1/p} .$$

$\|X+Y\|_p \le \|X\|_p + \|Y\|_p$

pf: (HW)

- $\displaystyle \sum_{i=1}^{n} |a_i b_i| \le \left(\sum_1^n a_i^p \right)^{1/p} \cdot \left(\sum_1^n b_i^q \right)^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1.$
- $\displaystyle \frac{1}{n} \left(\sum_1^n |a_i| \right)^2 \le \sum_1^n a_i^2 .$

Defn : A function $g(x)$ is convex [concave] if $g(\lambda x + (1-\lambda) y) \le [\ge] \lambda g(x) + (1-\lambda) g(y)$ for all x and y, $0 < \lambda < 1$.

Th^m (Jensen's ineq.) If g is a convex func., then

$$E\, g(X) \ge g(EX).$$

Equality holds iff for every line $a+bx$ that is tangent to $g(x)$ at EX

$$P(g(X) = a + bX) = 1.$$

(e.g.) arithmetic mean $\ge$ geometric mean $\ge$ harmonic mean.