Recall from last time:

<u>Markov's inequality</u>.

if X is a r.v. s.t. $X \geq 0$ a.s.,

then $P(X \geq a) \leq \frac{1}{a} E[X]$.

<u>Corollary</u>. (Chebyshev's inequality)

Let X be a r.v. s.t. $E[X]$ is finite and $Var(X) < \infty$

Then for any $a > 0$ we have

$$P(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2}.$$

<u>Proof</u>. $P(|X - E[X]| \geq a)$

$= P(|X - E[X]|^2 \geq a^2)$

$\leq \frac{1}{a^2} E[|X - EX|^2]$ by Markov $= \frac{1}{a^2} Var(X)$

<u>Application</u>. Weak law of large numbers.

Let $\xi_1, \dots, \xi_n$ are iid r.v.s with mean μ. s.t. $Var(\xi_i) < \infty$.

Let $S_N = \frac{1}{N} \sum_{i=1}^{N} \xi_i$. Then $\forall a > 0$.

$$P(|S_N - \mu| \geq a) \leq \frac{Var(\xi_1)}{Na^2}.$$

(note: RHS $\to 0$ as $N \to \infty$)

<u>Remark</u>. the condition that $Var(\xi_i) < \infty$ is unnecessary.

<u>Proof</u>. $Var(S_N) = \frac{1}{N^2} \sum_{i=1}^{N} Var(\xi_i) = \frac{1}{N} Var(\xi_1)$.

$E[S_N] = \frac{1}{N} \sum_{i=1}^{N} E[\xi_i] = \mu$.

Then Chebyshev $\Rightarrow P(|S_N - \mu| \geq a) \leq \frac{Var(S_N)}{a^2}$ " $= \frac{Var(\xi_1)}{Na^2}$

Random vectors

A random vector $\underline{X} = (X_1, \ldots, X_d)$ is just d random variables $X_1, \ldots, X_d$ on the same probability space.

Equivalently, $\underline{X}$ is a function from Ω to $\mathbb{R}^d$.

The distribution of a component of a random vector is called a marginal.

$P(X=0, Y=0) = 0.1$

$p_{X,Y}(x,y)$	$x=0$	$x=1$	p_Y
$y=0$	0.1	0.3	0.4
$y=1$	0.4	0.2	0.6
p_X	0.5	0.5	1

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

"marginalizing"

The joint cdf of a r.vec. $\underline{X}$ is $F_{\underline{X}}(x_1, \ldots, x_d) = P(X_1 \le x_1, \ldots, X_d \le x_d)$.

The joint pmf of $\underline{X}$ is $p_{\underline{X}}(x_1, \ldots, x_d) = P(X_1 = x_1, \ldots, X_d = x_d)$.

Theorem. The r.v.s X & Y are independent iff there are fns g and h s.t.

$$P_{X,Y}(x,y) = g(x)h(y).$$

More generally, $X_1, \dots, X_d$ are independent iff $\exists\, g_1, \dots, g_d$ s.t.

$$P_{\underline{X}}(x_1, \dots, x_d) = g_1(x_1) \cdots g_d(x_d).$$

Proof. (first statement $d=2$)

($\Rightarrow$) If X and Y are independent then just take $g = P_X$, $h = P_Y$.

($\Leftarrow$) Suppose $P_{X,Y}(x,y) = g(x)h(y)$.

Then $P_X(x) = \sum_y P_{X,Y}(x,y) = g(x) \sum_y h(y)$

Similarly,

$$P_Y(y) = h(y) \sum_x g(x).$$

So:

$$\boxed{\left(\sum_x g(x)\right)\left(\sum_y h(y)\right) = \sum_x \left(g(x) \sum_y h(y)\right)}$$

$$P_X(x) P_Y(y) = g(x) h(y) \sum_{x',y'} g(x') h(y')$$

$$= P_{X,Y}(x,y) \underbrace{\sum_{x',y'} P_{X,Y}(x',y')}_{\substack{\shortparallel \\ 1}}$$

$$= P_{X,Y}(x,y).$$ ∎

Note: g and h are constant multiples of the marginal distns.

Example. Suppose (X, Y) has joint pmf

$$p(x,y) = \frac{C \cdot 2^{-x}}{y}, \quad x = 1, 2, \ldots \quad y = 1, 2, \ldots, x.$$

& 0 otherwise.

where C is chosen so the probabilities sum to 1. i.e. $C^{-1} = \sum_{x,y : y \le x} \frac{2^{-x}}{y}$

Are X and Y independent?

NO.

Indeed,

$p(2, 100) = 0$

$\neq p_X(2) p_Y(100).$

a better way of writing p is

this term cannot be factored

$$p(x,y) = \frac{C \cdot 2^{-x}}{y} \mathbb{1}\{x \in \mathbb{N}\} \mathbb{1}\{y \in \mathbb{N}\} \mathbb{1}\{y \le x\}$$

Theorem Let X and Y be a random vector on a sample space, and let $Z = X + Y$.

Then the pmf of Z is

$$p_Z(z) = \sum_x p_{X,Y}(x, z-x) = \sum_y p_{X,Y}(z-y, y).$$

If X and Y are independent, then

$$p_Z(z) = \underbrace{\sum_x p_X(x) p_Y(z-x)}.$$

is called the convolution of p_X & p_Y.

Example. Suppose $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are indep. What is the distn of $Z = X+Y$?

$$p_Z(z) = \sum_{k=0}^{z} \frac{\lambda^k e^{-\lambda}}{k!} \cdot \frac{\mu^{z-k} e^{-\mu}}{(z-k)!}$$

$$= \frac{e^{-(\lambda+\mu)}}{z!} \sum_{k=0}^{z} \binom{z}{k} \lambda^k \mu^{z-k}$$

$$= \frac{e^{-(\lambda+\mu)}}{z!} (\lambda+\mu)^z.$$

So $Z \sim \text{Poisson}(\lambda+\mu)$.

Conditional pmfs

Let X & Y be r.v.s on Ω

The conditional pmf of Y given X is

$$p_{Y|X}(y|x) = P(Y=y \mid X=x)$$

The conditional cdf is

$$F_{Y|X}(y|x) = P(Y \leq y \mid X=x).$$

These are defined whenever $P(X=x) > 0$.

Thus, $p_{Y|X}(y|x) = \dfrac{p_{X,Y}(x,y)}{p_X(x)}$.

Let $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$ be independent; $Z = X+Y$.

Let's try to compute $p_{X|Z}$.

$$P(X=k \mid X+Y=m+k)$$

$$= \frac{P(X=k, Y=m)}{P(X+Y=k+m)} = \frac{P(X=k)P(Y=m)}{P(Z=k+m)}$$

$$= \frac{\dfrac{\cancel{e^{-\lambda}}}{k!}\lambda^k \cdot \dfrac{\cancel{e^{-\mu}}\mu^m}{m!}}{\dfrac{\cancel{e^{-(\lambda+\mu)}}(\lambda+\mu)^{m+k}}{(m+k)!}}$$

$$= \binom{m+k}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(1-\frac{\lambda}{\lambda+\mu}\right)^m.$$

So $p_{X|Z}(k|r) = \binom{r}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(1-\frac{\lambda}{\lambda+\mu}\right)^{r-k}$

So $X \mid Z=r \sim \text{Bin}\left(r, \frac{\lambda}{\lambda+\mu}\right)$

Conditional expectation.

$$E[Y|X=x] = \sum_y y P_{Y|X}(y|x).$$

So in the last example,

$$E[X|X+Y=r] = \frac{r\lambda}{\lambda+\mu}.$$

We can also think of

$E[Y|X]$ as a random variable itself. For $\omega \in \Omega$, we define

$$E[Y|X](\omega) = E[Y|X=X(\omega)]$$
$$= \sum_y y P_{Y|X}(y|X(\omega)).$$

Note: $E[Y]$ is deterministic. But $E[Y|X]$ is (in general) random.

in the last example

$$E[X|X+Y] = \frac{(X+Y)\lambda}{\lambda+\mu}.$$

Theorem

$$E[Y] = \sum_x E[Y|X=x]P(X=x)$$
$$= E[E[Y|X]].$$

Theorem

$$E[Y] = \sum_x E[Y \mid X = x] P(X = x)$$
$$= E\left[E[Y \mid X]\right].$$

Proof

$$E[Y] = \sum_y y P(Y = y)$$
$$= \sum_y y \sum_x P(Y = y, X = x)$$
$$= \sum_y y \sum_x p_{Y|X}(y|x)\, p_X(x)$$
$$= \sum_x p_X(x) \sum_y y\, p_{Y|X}(y|x)$$
$$= \sum_x p_X(x)\, E[Y \mid X = x].$$
$$= E[E[Y \mid X]]$$