- (1.3.8) $E(X|k) = E(X|Y=k) = \sum_{j=0}^{\infty} j\, P_{X|Y}(j|k)$

provided $P_Y(k) > 0$, and is undefined otherwise.

- We can view $E(X|Y=y)$ as a function of y which equals $E(X|Y=k)$ when $y=k$.

Definition 1.3.7

We write $E(X|Y)$ for this function and we call it the conditional expectation of X given Y.

— #3 1/29/08

Example 1.3.5

Let X have a uniform distribution on $0, 1, 2, \ldots, n$ and given $X=j$, let Y have a uniform distribution on $0, 1, 2, \ldots, j$.

We have $Y | X = j \sim \text{unif}\{0, 1, 2, \ldots, j\}$, so

$$E(Y|X=j) = \frac{j}{2}$$

This means that

$$E(Y|X) = \frac{X}{2}$$

Theorem 1.3.9 Law of Total Probability

Let X, Y be integer valued, nonnegative random variables. Then

(1.3.9) $$E(E(X|Y)) = E(X).$$

Proof

Let f_Y = p.m.f. of Y.

$$E(E(X|Y)) = \sum_{k=0}^{\infty} E(X|k) f_Y(k)$$

$$= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} j \, P_{X|Y}(j|k) f_Y(k)$$

$$= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} j \, P_{X,Y}(j,k)$$

$$= \sum_{j=0}^{\infty} j \, P_X(j)$$

$$= E(X).$$

—

We can interpret this theorem as saying

(1.3.10) $$E(X) = \sum_{k=0}^{\infty} E(X|k) P(Y=k)$$

Example 1.3.6

A hen lays N eggs where $N \sim$ pois(λ). Each egg hatches with probability p independently of the other eggs. Let K be the number of chicks.
Compute $E(K|N)$, $E(K)$, $E(N|K)$.

The pmf for N is

$$f_N(n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

and the cond. p.m.f.

$$P_{K|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}$$

So

$$E(K|n) = \sum_{k=0}^{n} k \, P_{K|N}(k|n)$$

$$= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = pn$$

Thus, $E(K|N) = pN$ and

$$E(K) = E(E(K|N)) = pE(N) = p\lambda$$

To compute $E(N|K)$, we use $P_{N|K}$.

$$p_{N|K}(n|k) = P(N=n \mid K=k) \quad \downarrow \text{using } P(N=n, K=k) = P(K=k, N=n)$$

$$= \frac{P(K=k \mid N=n)\, P(N=n)}{P(K=k)}$$

$$= \frac{\binom{n}{k} p^k (1-p)^{n-k} \, (\lambda^n/n!)\, e^{-\lambda}}{\sum_{m \ge k} \binom{m}{k} p^k (1-p)^{m-k} \, (\lambda^m/m!)\, e^{-\lambda}} \qquad n \ge k$$

$\downarrow$ tedious

$$= \frac{(q\lambda)^{n-k}}{(n-k)!} e^{-q\lambda}$$

Hence,

$$E(N \mid K=k) = \sum_{n \ge k} n \frac{(q\lambda)^{n-k}}{(n-k)!} e^{-q\lambda} = k + q\lambda$$

and

$$E(N \mid K) = K + q\lambda.$$

There is a more general version of Thm 1.3.9.

<u>Definition 1.3.8</u>

Let X, Y be integer valued, non negative random variables. Let g be a function such that $g(X)$ is finite. The <u>conditional expected value</u> of $g(X)$ given $Y = k$ is

$$E(g(X) \mid k) = \sum_{j=0}^{\infty} g(j)\, p_{X|Y}(j|k)$$

when $P_{Y}(k) > 0$ and undefined otherwise.

Theorem 1.3.10

Let X, Y be integervalued, nonnegative random variables and g a function such that $E(g(X)) < \infty$. Then,

$$(1.3.11) \quad E(g(X)) = E\Big(E(g(X) \mid Y)\Big).$$

Proof

Exercise.

(Please see page 61 in the text for important properties of conditional expected values.)

§1.4 Convolution

We often deal with sums of random variables, for which we want to compute information.

Let X, Y be independent, nonnegative integer valued random variables with

$$X \sim \{a_k\} \iff P(X = k) = a_k$$

$$Y \sim \{b_k\}$$

Since X and Y are independent

$$P(X+Y=n) = \sum_{k=0}^{n} P(X=k, Y=n-k)$$

$$= \sum_{k=0}^{n} P(X=k)\,P(Y=n-k)$$

$$= \sum_{k=0}^{n} a_k b_{n-k} = p_n.$$

In general, we define

Definition 1.4.1

The convolution of two sequences $\{a_n\}_{n=0}^{\infty}$ $\{b_n\}_{n=0}^{\infty}$ is the sequence $\{c_n\}_{n=0}^{\infty}$ with

$$c_n = \sum_{i=0}^{n} a_i b_{n-i}$$

We write

$$\{c_n\} = \{a_n\} * \{b_n\}$$

We proved above

Theorem 1.4.1

If X, Y are independent, integer valued nonnegative random variables

$$\{P(X+Y=n)\} = \{P(X=n)\} * \{P(Y=n)\}$$

Example 1.4.1

Suppose $X \sim Pois(k,\lambda)$, $Y \sim Pois(k,\mu)$.

$$P(X+Y=n) = \sum_{k=0}^{n} Pois(k,\lambda)\, Pois(n-k,\lambda)$$

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

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

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

These are the basic properties of convolution

Theorem 14.2

Suppose $\{a_j\}$, $\{b_j\}$, $\{d_j\}$ are probability mass functions.

(1) $\{c_j\} = \{a_j\} * \{b_j\}$ is a probability mass function

(2) $\{a_j\} * \{b_j\} = \{b_j\} * \{a_j\}$

(If $X \sim \{a_j\}$, $Y \sim \{b_j\}$, we write $X + Y \overset{d}{=} Y + X$, where "$\overset{d}{=}$" means "has the same distribution as".)

(3) $(\{a_j\} * \{b_j\}) * \{d_j\} = \{a_j\} * (\{b_j\} * \{d_j\})$

(4) If $X_1, \ldots, X_k$ are iid (independent, identically distributed) with $X_1 \sim \{a_j\}$, then

$$X_1 + \cdots + X_k \sim \underbrace{\{a_j\} * \cdots * \{a_j\}}_{k \text{ copies}}$$

Proof

Exercise

Definition 1.4.2

We write

$$\{a_n\}^{k*} = \underbrace{\{a_n\} * \cdots * \{a_n\}}_{k \text{ copies}}$$

§1.5 Generating Functions

See TK Ch. III, §9

We describe a way to store the information in a sequence using a single function.

Definition 1.5.1

Let $\{a_j\}_{j=0}^{\infty}$ be a sequence. If there is an $s_0 > 0$ such that

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

converges for $|s| < s_0$, we call $A(s)$ the probability generating function (pgf) of $\{a_j\}$.

As motivation, consider that formally,

$$a_j = \frac{1}{j!} \frac{d^j}{ds^j} A(s) \Big|_{s=0}.$$

There is more than one way to define a pgf, but Defn. 1.5.1 is particularly useful when $\{a_j\}$ is a pmf.

Definition 1.5.2

Let X be a non negative, integer valued random variable with $X \sim \{p_k\}_{k=0}^{\infty}$. The

<u>generating function</u> of X is

$$P(s) = \sum_{k=0}^{\infty} p_k s^k$$

(which is the pgf of $\{p_k\}$).

Note that

$$P(s) = E(s^X) \quad , \; 0 \le s \le 1$$

Also note that

$$P(1) = \sum_{k=1}^{\infty} p_k \le 1,$$

so the radius of convergence for $P(s)$ is at least 1 (it may be larger).

<u>Example 1.5.1</u>

Let X have binomial distribution with parameters n, p.

$$E(s^X) = \sum_{k=0}^{n} s^k \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (sp)^k (1-p)^{n-k}$$

$$= (1-p+sp)^n$$

Example 1.5.2

If $X \sim Pois(k, \lambda)$,

$$E(s^X) = \sum_{k=0}^{\infty} s^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}$$

Returning to the justification for the definition, since $P(s)$ converges absolutely at least for $|s| \leq 1$, we also know that $P(s)$ is infinitely differentiable at least for $|s| < 1$, and we can differentiate term by term in the series.

We find

$$\frac{d^n}{ds^n} P(s) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1) p_k s^{k-n}, \quad |s| < 1,$$

Setting $s=0$ yields

Theorem 1.5.1

Let $X \sim \{p_k\}$ be a nonnegative, integer valued random variable. Then

$$p_n = \frac{1}{n!} \frac{d^n}{ds^n} P(s) \Big|_{s=0}$$

Theorem 1.5.2

A pmf $\{p_k\}$ is uniquely determined by its pgf.

There is an interesting connection between convolution and generating functions that can be used to compute convolutions.

Theorem 1.5.3

The pgf of a convolution of pmf's is the product of the pgf's of the pmf's.

(1) If X, Y are independent, non-negative integer valued random variables with pgf's

$$P_X(s) = E(s^X), \quad P_Y(s) = E(s^Y) \quad , 0 \le s \le 1,$$

then

$$P_{X+Y}(s) = P_X(s) P_Y(s).$$

(2) If $\{a_j\}, \{b_j\}$ are sequences with

pgf's $A(s)$, $B(s)$, then the pgf of $\{a_j\} * \{b_j\}$ is $A(s)B(s)$.

Note: If $X \stackrel{d}{=} Y$, then $P_{X+Y}(s) = (P_X(s))^2$.

<u>Proof</u>

(2) => (1), but we prove each.

(1) $$P_{X+Y}(s) = E(s^{X+Y})$$
$$= E(s^X s^Y)$$
$$= E(s^X)\,E(s^Y)$$
$$= P_X(s)\,P_Y(s).$$

(2) Without loss of generality, let the radius of convergence of $A(s)$ and $B(s)$ be s_0. The pgf of the convolution is

$$\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} a_k b_{n-k}\right)s^n, \quad |s| < s_0.$$

Fubini's Theorem allows us to switch the order of summation

$$\sum_{k=0}^{\infty} \sum_{n=k}^{\infty} a_k b_{n-k} s^n$$

$$= \sum_{k=0}^{\infty} a_k s^k \sum_{n=k}^{\infty} b_{n-k} s^{n-k}$$

$$= A(s) B(s).$$

Example 1.5.3

If X, Y are independent, $X \sim Pois(\lambda)$, $Y \sim Pois(\mu)$, then

$$P_{X+Y}(s) = P_X(s) P_Y(s) = e^{\lambda(s-1)} e^{\mu(s-1)}$$
$$= e^{(\lambda+\mu)(s-1)}.$$

We conclude that $X + Y \sim Pois(\lambda+\mu)$.

$\boxed{\text{Start \#4} \quad 1/31/08}$

Example 1.5.4

Let X be the number of failures necessary to obtain r successes in repeated independent Bernoulli trials. The density of X is called the negative binomial distribution. We can represent X as a sum. We let $\{X_1, \dots, X_r)$ be iid rv with geometric distribution