Discrete Random Variables
lec05-1

Dr Ahmad Khonsari

ECE Dept.

The University of Tehran

Synopsis of transforms

In problems where random variables are nonnegative, it is usually more convenient to use the z-transform or the Laplace transform. The **probability generating function** $G_N(z)$ of a nonnegative integer-valued random variable N is defined by

$$G_N(z) = E[z^N] \tag{4.84a}$$

$$= \sum_{k=0}^{\infty} p_N(k) z^k. \tag{4.84b}$$

The first expression is the expected value of the function of N, z^N. The second expression is the z-transform of the pmf (with a sign change in the exponent).

In queueing theory one deals with service times, waiting times, and delays. All of these are nonnegative continuous random variables. It is therefore customary to work with the **Laplace transform** of the pdf,

$$X^*(s) = \int_0^{\infty} f_X(x) e^{-sx}\, dx = E[e^{-sX}]. \tag{4.88}$$

Note that $X^*(s)$ can be interpreted as a Laplace transform of the pdf or as an expected value of a function of X, e^{-sX}.

Calculating the moments of the distribution with the aid of $G(z)=\sum_{i=0}^{\infty} p_i z^i$

Note: Since the p_i represent a probability distribution their sum equals 1 and

$$G(1) = G^{(0)}(1) = \sum_{i=0}^{\infty} p_i 1^i = 1$$

By derivation one sees

$$G^{(1)}(z) = \frac{d}{dz} E[z^X] = E[X z^{X-1}]$$

$$G^{(1)}(1) = E[X]$$

By continuing in the same way one gets

$$\boxed{G^{(i)}(1) = E[X(X-1)\cdots(X-i+1)] = F_i}$$

where F_i is the i^{th} factorial moment.

The relation between factorial moments and ordinary moments (with respect to the origin)

The factorial moments $F_i = \mathrm{E}[X(X-1)\cdots(X-i+1)]$ and ordinary moments (with respect to the origin) $M_i = \mathrm{E}[X^i]$ are related by the linear equations:

$$\begin{aligned} M_1 &= F_1 \\ M_2 &= F_2 + F_1 \\ M_3 &= F_3 + 3F_2 + F_1 \\ &\vdots \end{aligned} \qquad\qquad \begin{aligned} F_1 &= M_1 \\ F_2 &= M_2 - M_1 \\ F_3 &= M_3 - 3M_2 + 2M_1 \end{aligned}$$

For instance,

$$F_1 = G^{(1)}(1) = G^{(1)}(z)\big|_{z=1} = \frac{d}{dz}E[z^X]\Big|_{z=1} = E[Xz^{X-1}]|_{z=1} = \mathrm{E}[X] \qquad M_1 = \mathrm{E}[X]$$

$$F_2 = G^{(2)}(1) = \mathrm{E}[X(X-1)] = \mathrm{E}[X^2] - \mathrm{E}[X]$$

$$\Rightarrow \quad M_2 = \mathrm{E}[X^2] = F_2 + F_1 = G^{(2)}(1) + G^{(1)}(1) = \mathrm{E}[X^2] - \mathrm{E}[X] + \mathrm{E}[X]$$

$$\Rightarrow \quad \mathrm{V}[X] = M_2 - M_1^2 = G^{(2)}(1) + G^{(1)}(1) - (G^{(1)}(1))^2 = G^{(2)}(1) + G^{(1)}(1)(1 - G^{(1)}(1))$$

Direct calculation of the moments

The moments can also be derived from the generating function directly, without recourse to the factorial moments, as follows:

$$\frac{d}{dz}G(z)\bigg|_{z=1} = E[XZ^{X-1}]_{z=1} = E[X]$$

$$\frac{d}{dz}z\frac{d}{dz}G(z)\bigg|_{z=1} = \frac{d}{dz}zE[XZ^{X-1}]\bigg|_{z=1} = \frac{d}{dz}E[XZ^{X}]\bigg|_{z=1} = E[X^2Z^{X-1}]_{z=1} = E[X^2]$$

Generally,

$$\frac{d}{dz}\left(z\frac{d}{dz}\right)^{i-1}G(z)\bigg|_{z=1} = E\left[X^i\right]$$

Generating function of the sum of independent random variables

Let X and Y be ***independent*** random variables. Then

$$G_{X+Y}(z) = \mathrm{E}[z^{X+Y}] = \mathrm{E}[z^X z^Y]$$

$$= \mathrm{E}[z^X]\mathrm{E}[z^Y] \qquad \text{independence}$$

$$= G_X(z)G_Y(z)$$

$$G_{X+Y}(z) = \boxed{G_X(z)G_Y(z)}$$

In terms of the original discrete distributions

$p_i = \mathrm{P}\{X = i\}$

$q_j = \mathrm{P}\{Y = j\}$

the distribution of the sum of ***independent*** RVs is obtained by convolution $p \odot q$

$$\mathrm{P}\{X + Y = k\} = (p \odot q)_k = \sum_{i=0}^{k} p_i \, q_{k-i}$$

Thus, the generating function of a distribution obtained by convolving two independent distributions is the product of the generating functions of the respective original distributions.

Q: what is the distribution of the ***difference*** of 2 independent RVs ?

Random SUM of iid RVS (Compound distribution) :

Let S_N be the sum of independent, identically distributed (*i.i.d.*) random variables X_i, with common mean and variance

$S_N = X_1 + X_2 + \cdots X_N$ where N is a non-negative integer-valued RV also independent from all X_i,. What is E[S_N], var[S_N] and distribution of S_N

E[S_N]: for fixed n

$$\mathbf{E}\left[S_N \mid N = n\right] = \mathbf{E}\left[X_1 + X_2 + \cdots + X_N \mid N = n\right]$$
$$= \mathbf{E}\left[X_1 + X_2 + \cdots + X_n \mid N = n\right] = \mathbf{E}\left[X_1 + X_2 + \cdots + X_n\right] = n\mathbf{E}\left[X_i\right]$$

$\mathrm{E}[S_N|N]$ can be viewed as a function of random N variable

$\mathrm{E}[S_N|N]$ is a random variable

$\mathrm{E}[S_N] = \mathrm{E}[\mathrm{E}[S_N \mid N]]$ the law of iterated expectations

$=\mathrm{E}[N\mathrm{E}[X]]$

$=\mathrm{E}[N]\mathrm{E}[X]$

Aside: conditional expectation

$\mathrm{E}[\mathrm{Y}|\mathrm{X} = x] = \int_{-\infty}^{\infty} y f_{\mathrm{Y}|X}(y|x)\, dy$ $\qquad E[\mathrm{Y}|\mathrm{X} = x] = \sum_{\forall y} y p_{\mathrm{Y}|X}(y|x)$

$$\mathrm{E}[Y] = \underset{x\ \ y}{\mathrm{E}[\mathrm{E}[\mathrm{Y}|\mathrm{X} = x]]} = \int_{-\infty}^{\infty} \mathrm{E}[\mathrm{Y}|\mathrm{X} = x] f_X(x)\, dx$$ the law of iterated expectations

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{\mathrm{Y}|X}(y|x)\, dy\, f_X(x)\, dx = \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{\mathrm{X,Y}}(x,y) dx dy = \int_{-\infty}^{\infty} y f_{\mathrm{Y}}(y) dy$$

Random SUM of iid RVS (Compound distribution) : Variance

var [S_N]

- for fixed n, $\mathrm{var}[S_N|N=n]$ can be expressed as

$$\begin{aligned}
&\mathrm{var}\left(S_N \mid N=n\right)\\
&=\mathrm{var}\left(X_1+X_2+\cdots+X_N \mid N=n\right)\\
&=\mathrm{var}\left(X_1+X_2+\cdots+X_n \mid N=n\right)\\
&=\mathrm{var}\left(X_1+X_2+\cdots+X_n\right)\\
&=n\,\mathrm{var}\left(X\right)
\end{aligned}$$

independence

  – $\mathrm{var}[S_N|N]$ can be viewed as a function of random variable N
    - $\mathrm{var}[S_N|N]$ is a random variable
  – The variance of S_N can be calculated using the law of total variance

$$\begin{aligned}
\mathrm{var}\left(S_N\right) &= \mathbf{E}\left[\mathrm{var}\left(S_N \mid N\right)\right]+\mathrm{var}\left(\mathbf{E}\left[S_N \mid N\right]\right)\\
&= \mathbf{E}\left[N\,\mathrm{var}\left(X\right)\right]+\mathrm{var}\left(N\,\mathrm{E}[\mathrm{X}]\right)\\
&= \mathrm{var}\left(X\right)\mathbf{E}\left[N\right]+\left(\mathbf{E}\left[X\right]\right)^2\mathrm{var}\left(N\right)
\end{aligned}$$

Random SUM of iid RVS (Compound distribution) :

Denote:

$G_X(z)$ the common generating function of the X_i

$G_N(z)$ the generating function of N

For fixed n:

$$E[z^{S_N}|N = n] = [G_X(z)]^n \qquad n = 1, 2, 3, \ldots .$$

Proof:

$$\begin{aligned} E[z^{S_N}|N = n] &= E[z^{X_1+X_2+\cdots+X_N}|N = n] \\ &= E[z^{X_1} z^{X_2} \ldots z^{X_N}|N = n] \\ &= E[z^{X_1} z^{X_2} \ldots z^{X_n}] \\ &= E[z^{X_1}] E[z^{X_2}] \ldots E[z^{X_n}] \qquad \text{independence} \\ &= [G_X(z)]^n \end{aligned}$$

We wish to calculate $G_{S_N}(\mathrm{z})$

The transformation $G_{S_N}(\mathrm{z})$ is found by starting with $G_N(\mathrm{z})= E[z^N]$ transform and replacing each occurrence of z with $G_X(\mathrm{z})$ i.e.

$$G_N(\mathrm{z})\Big|_{G_X(Z)} = E[Z^N]\Big|_{G_X(Z)}$$

Prove that $G_{S_N}(\mathrm{z}) = G_N\left(G_X(z)\right)$

Proof:

$$\begin{aligned} G_{S_N}(\mathrm{z}) &= \mathrm{E}[z^{S_N}] \\ &= \mathrm{E}[\mathrm{E}[z^{S_N}|\mathrm{N}]] \\ &= \mathrm{E}[G_X(z)^N] \\ &= G_N\left(G_X(z)\right) \end{aligned}$$

Law of iterated expectation

Using $\mathrm{E}[z^{S_N}|\mathrm{N} = n] = [G_X(z)]^n$

Analogy

$G_N(\mathrm{z}) = E[z^N] \leftrightarrow G_N(G_X(z))=\mathrm{E}[G_X(z)^N]$

فرض: $\{X_i\}$ ها iid ، $N \perp X_i$ ، $S=\sum_{i=1}^{N} X_i$ داریم:

$$E[S]=E[N]E[X]$$

$$E[S^2]=E[N]\mathrm{Var}(X)+E[N^2](E[X])^2$$

$$\mathrm{Var}(S)=E[N]\mathrm{Var}(X)+\mathrm{Var}(N)(E[X])^2$$

reminder:

واریانس شرطی :

$$\mathrm{Var}(X|Y=y)=E\left[(X-E[X|Y=y])^2 \mid Y=y\right]$$

قانون واریانس کل :

$$\mathrm{Var}(X)=\underbrace{E(\mathrm{Var}(X|Y))}_{\text{unexplained}}+\mathrm{Var}\Big(\underbrace{E(X|Y)}_{\substack{\text{متغیر تصادفی که فقط به } Y \text{ ربط دارد} \\ \text{explained}}}\Big)$$

فرض کنید $\{X_i\}$ ها iid باشند و N تعدادی از $\{X_i\}$ ها باشد که متغیر تصادفی مستقل از $\{X_i\}$ ها باشد. فرض کنید

$$S=\sum_{i=1}^{N} X_i \qquad \overset{\text{علامت استقلال}}{N \perp X_i}$$

$E(S)$ و $E(S^2)$ را حساب کنید؟

قانون خطی بودن امید برای متغیر تصادفی مستقیماً قابل اعمال نمی‌باشد. ولی برای n ثابت قابل اعمال است.

$$n \in support\ \mathrm{N}$$

$$E(S)=E\left[\sum_{i=1}^{N} X_i\right]=\sum_{n} E\left[\sum_{i=1}^{N} X_i \mid N=n\right]\cdot P(N=n)$$

$$=\sum_{n} E\left[\sum_{i=1}^{n} X_i\right]\cdot P(N=n) \quad \text{:}$$

طبق $\overset{\text{استقلال}}{N \perp X_i}$ و

$$=\sum_{n} n\, E(X)\cdot P(N=n)$$

iid بودن $\{X_i\}$: $E[\sum X_i]=\sum E(X)$

$$=E(X)\sum n\,P(N=n)$$

$$=E(X)E(N)$$

آیا می‌توان شگرد فوق را برای محاسبه $E(S^2)$ بکار برد؟ خیر چون در بین راه مواجه با یک مجموع تصادفی به صورت $\sum_{i=1}^{N} X_i$ خواهیم بود که باید به توان ۲ رساند (یعنی محاسبه $\left(\sum_{i=1}^{N} X_i\right)^2$ مشکل است)

$$E(S^2) = \sum_n E(S^2 | N=n)\, P(N=n)$$

$$= \sum_n E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \cdot P(N=n)$$

که محاسبه‌اش بسیار مشکل است.

روش بهتر؟ خوب است ابتدا $Var(S|N=n)$ را بیابیم و از روی آن $E(S^2|N=n)$ را محاسبه کنیم.

می‌دانیم

$$Var(S|N=n) = n\,Var(X)$$

$$n\,Var(X) = Var(S|N=n) = E[S^2 | N=n] - \left(E[S|N=n]\right)^2$$

$$= E[S^2|N=n] - \left(n E[X]\right)^2$$

از اینجا داریم:

$$E[S^2|N=n] = n\,Var(X) + n^2 (E[X])^2$$

حال داریم:

$$E[S^2] = \sum_n E[S^2 \mid N=n] \cdot P(N=n)$$

$$= \sum_n \left(n\,Var(X) + n^2 E(X)^2 \right) P(N=n)$$

$$= E[N]\,Var(X) + E[N^2]\left(E(X)\right)^2$$

در ضمن $Var(S)$ هم بدست آوریم:

$$Var(S) = E[S^2] - \left(E[S]\right)^2$$

$$= E[N]\,Var(X) + \underline{E[N^2]\left(E[X]\right)^2 - \left(E[N]E(X)\right)^2}$$

$$= E[N]\,Var(X) + \left\{ \underline{E[N^2] - E[N]^2} \right\}\left(E[X]\right)^2$$

$$= E[N]\,Var(X) + Var(N)\left(E[X]\right)^2$$

The distribution of max and min of independent RVs

Let $X_1, X_2, \ldots, X_n$ be independent random variables (distribution functions $F_i(x)$ and tail distributions $G_i(x)$, $i = 1, \ldots, n$)

Distribution of the maximum

$P\{\max(X_1, X_2, \ldots, X_n) \le x\} = P\{X_1 \le x, \ldots, X_n \le x\}$

$= P\{X_1 \le x\} \cdots P\{X_n \le x\}$ (independence!)

$= F_1(x) \cdots F_n(x)$

$= (F(x))^n$ iid

Distribution of the minimum

$P\{\min(X_1, X_2, \ldots, X_n) > x\} = P\{X_1 > x, \ldots, X_n > x\}$

$= P\{X_1 > x\} \cdots P\{X_n > x\}$ (independence!)

$= G_1(x) \cdots G_n(x)$

$= (G(x))^n$ iid

The distribution of max and min of independent RVs

Theorem 5.7 *Let $\mathbf{X}$ be a vector of n iid random variables each with CDF $F_X(x)$ and PDF $f_X(x)$.*

(a) The CDF and the PDF of $Y = \max\{X_1, \ldots, X_n\}$ are

$$F_Y(y) = (F_X(y))^n, \qquad f_Y(y) = n(F_X(y))^{n-1} f_X(y).$$

(b) The CDF and the PDF of $W = \min\{X_1, \ldots, X_n\}$ are

$$F_W(w) = 1 - (1 - F_X(w))^n, \qquad f_W(w) = n(1 - F_X(w))^{n-1} f_X(w).$$

Proof By definition, $f_Y(y) = P[Y \le y]$. Because Y is the maximum value of $\{X_1, \ldots, X_n\}$, the event $\{Y \le y\} = \{X_1 \le y, X_2 \le y, \ldots, X_n \le y\}$. Because all the random variables X_i are iid, $\{Y \le y\}$ is the intersection of n independent events. Each of the events $\{X_i \le y\}$ has probability $F_X(y)$. The probability of the intersection is the product of the individual probabilities, which implies the first part of the theorem: $F_Y(y) = (F_X(y))^n$. The second part is the result of differentiating $F_Y(y)$ with respect to y. The derivations of $F_W(w)$ and $f_W(w)$ are similar. They begin with the observations that $F_W(w) = 1-P[W > w]$ and that the event $\{W > w\} = \{X_1 > w, X_2 > w, \ldots X_n > w\}$, which is the intersection of n independent events, each with probability $1 - F_X(w)$.

ORDER STATISTICS

Let $X_1, X_2, \ldots, X_n$ be mutually iid continuous RVs, each having the distribution function F and density f.

Let $Y_1, Y_2, \ldots, Y_n$ be a permutation of the set $X_1, X_2, \ldots, X_n$ so as to be in increasing order.

To be specific:

$Y_1 = \min\{X_1, X_2, \ldots, X_n\}$

and

$Y_n = \max\{X_1, X_2, \ldots, X_n\}$.

Y_k is called the **k**$^{\text{th}}$**-order statistic.**

Since $X_1, X_2, \ldots, X_n$ are continuous RVs, it follows that

$Y_1 < Y_2 <, \ldots < Y_n$ (as opposed to $Y_1 \leq Y_2 \leq \ldots \leq Y_n$) with a probability of one.

ORDER STATISTICS

As examples of use of order statistics,

let X_i; be the lifetime of the ith component in a system of n independent components.

e.g.

series system,

Y_1 is overall system lifetime of a series system.

Parallel systems

Y_n is the lifetime of a parallel system

and

Y_{n-k+1} is the lifetime of an k-out of-n system (the so-called N-tuple Modular Redundant or NMR system).

ORDER STATISTICS

Deriving the distribution function of Y_k

the probability that exactly

j of the X_i's lie in$(-\infty, y]$ and (n - j) lie in $(y, \infty)$is:

$\binom{n}{j} F^j(y)[l - F(y)]^{n-j}$ since the binomial distribution with parameters n and p = F(y) is applicable.

Then:

$F_{Y_k}(y) = p(Y_k \leq y) = P$ ("at least k of the X_i*'s* lie in the interval $(-\infty, y]$ ")

$= \sum_{j=k}^{n} \binom{n}{j} F^j(y)[l - F(y)]^{n-j} \qquad -\infty \leq y \leq \infty \qquad (3.52)$

ORDER STATISTICS

In particular, the distribution functions of Y_n and Y_1 (*i.e.* **max and min**) can be obtained from (3.52) as:

$$F_{Y_n}(y) = \sum_{j=n}^{n} \binom{n}{j} F^j(y)[l - F(y)]^{n-j} = [F(y)]^n \qquad -\infty \le y \le \infty \quad ,$$

$$\sum_{i=0}^{n} p_i = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = [p + (1-p)]^n = 1.$$

$$F_{Y_1}(y) = \sum_{j=1}^{n} \binom{n}{j} F^j(y)[l - F(y)]^{n-j}$$

$$= \sum_{j=0}^{n} \binom{n}{j} F^j(y)[l - F(y)]^{n-j} - \binom{n}{0} F^0(y)[l - F(y)]^{n-0} = 1 - [1 - F(y)]^n \qquad -\infty \le y \le \infty$$

Thus we obtain:

$R_{series}(t) = R_{Y_1}(t) = 1 - F_{Y_1}(t) = 1 - (1 - [1 - F(t)]^n) = [1 - F(t)]^n = [R(t)]^n$,

$R_{parallel}(t) = R_{Y_n}(t) = 1 - F_{Y_n}(t) = 1 - [F(t)]^n = 1 - [1 - R(t)]^n$

We may generalize above to the case when the lifetime distributions of individual components are distinct:

$R_{series}(t) = \prod_{i=1}^{n} R_i(t)$, $R_{parallel}(t) = 1 - \prod_{i=1}^{n} (1 - R_i(t))$

7. Important distributions

We will deal with:

- discrete distributions:
  - Bernoulli;
  - binomial;
  - geometric;
  - Negative Binomial;
  - Poisson.

7.1. The Bernoulli(p) $X \sim \text{Bernoulli}(p)$

Assume we have one experiment:

event A occurs with probability p; $\Pr[\{A\}]=p$ $0 \le p \le 1$

event A does not occur with probability $(1 - p)$; $\Pr[\{\overline{A}\}]=1-p=q$ $0 \le p \le 1$

$\Omega=\{A,\overline{A}\}$

If X is a r.v. drawn from the Bernoulli(p) distribution, write: $X \sim \text{Bernoulli}(p)$

and we define RV X as:

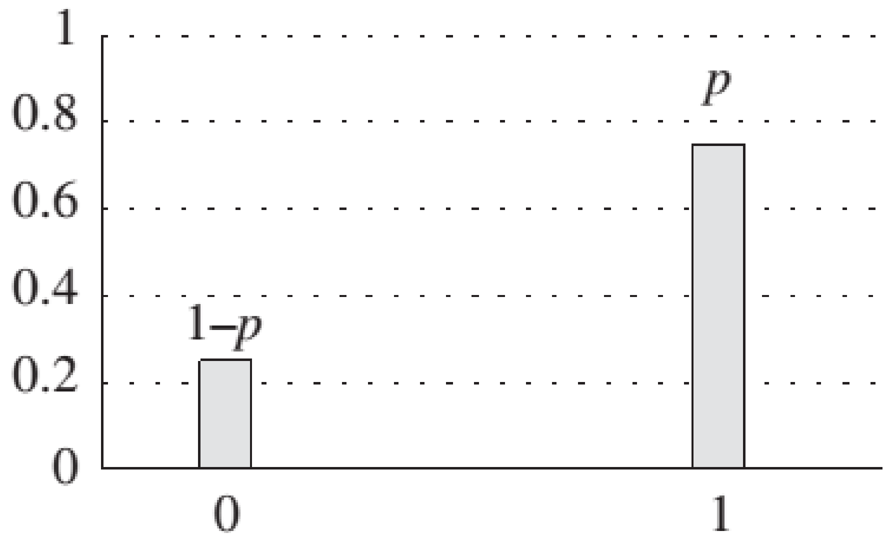
$$X = \begin{cases} 1 & \text{w/ prob } p \\ 0 & otherwise \end{cases}$$

The p.m.f. of r.v. X is defined as :

$P_X(1) = p$

$P_X(0) = 1 - p$

Example: Bernoulli(p), p=0.75



Mean and Variance of a Bernoulli Random Variable

The **mean** is:

$$\mu_X = E(X) = \sum_X xP(x) = (0)(1-P) + (1)P = P$$

And the **variance** is:

$$\sigma_X^2 = E[(X-\mu_X)^2] = \sum_X (x-\mu_X)^2 P(x)$$

$$= (0-P)^2(1-P) + (1-P)^2 P = P(1-P)$$

Bernoulli distribution $X \sim \mathrm{Bernoulli}(p)$

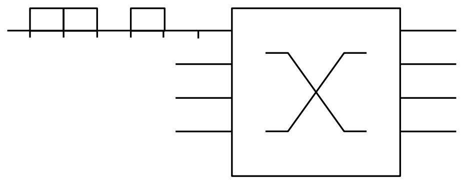
Example 1. X describes the bit stream from a traffic source, which is either on or off. The generating function

$G(z) = p_0 z^0 + p_1 z^1 = q + pz$

$E[X] = G^{(1)}(1) = p$

$V[X] = G^{(2)}(1) + G^{(1)}(1)(1 - G^{(1)}(1)) = p(1-p) = pq$

Example 2. The cell stream arriving at an input port of an ATM switch: in a time slot (cell slot) there is a cell with probability p or the slot is empty with probability q.

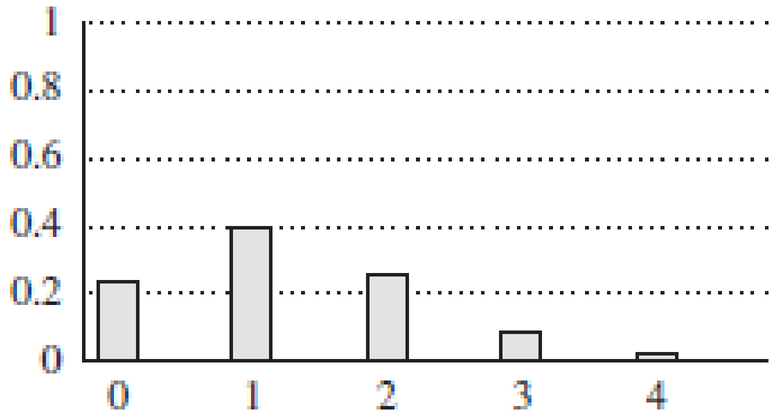


7.2. Binomial(n,p) distribution $X \sim \mathrm{Bin}(n, p)$

Definition: If $X \sim$ Binomial(n, p), then X represents the number of successes in n Bernoulli(p) experiments (i.e.
$X = \sum_{i=1}^{n} Y_i$ where $Y_i \sim$Bernoulli(p) and the Y_i are independent ($i = 1, \ldots, n$)

The p.m.f. of r.v. X is defined as follows

$$\Pr\{X = i\} = C_n^i p^i (1-p)^{n-i} \quad i = 0,1,\ldots,n,\ 0 \le p \le 1.\ C_n^i = \binom{n}{i}$$



e.g. Probability mass function of the Binomial(n, p) distribution, with n = 4 and p = 0.3.

Note: prove that Binomial is a probability distribution

$$\sum_{i=0}^{n} p_i = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$
$$= [p + (1-p)]^n$$
$$= 1.$$

Binomial distribution $X \sim \mathrm{Bin}(n,\ p)$

The generating function is obtained directly from the generating function $q + pz$ of a Bernoulli variable $G(z) = (q + pz)^n$ or directly from the definition

$$G(z) = \sum_{i=0}^{\infty} p_i z^i = \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i} z^i = \sum_{i=1}^{n} \binom{n}{i} (pz)^i (1-p)^{n-i} = (q+pz)^n$$

By identifying the coefficient of z^i in the expansion of we have

$$p_i = P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

$$\mathrm{E}[X] = n\mathrm{E}[Y_i] = np$$

$$\mathrm{V}[X] = n\mathrm{V}[Y_i] = np(1-p)$$

A limiting form when $\lambda = \mathrm{E}[X] = np$ is fixed and $n \to \infty$: using $e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$

$$G(z) = (1 - (1-z)p)^n = \left(1 - (1-z)\frac{\lambda}{n}\right)^n \to e^{(z-1)\lambda}$$

which is the generating function of a Poisson random variable.

Theorem 6.2 *The variance of $W_n = X_1 + \cdots + X_n$ is*

$$\operatorname{Var}[W_n] = \sum_{i=1}^{n} \operatorname{Var}[X_i] + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \operatorname{Cov}\left[X_i, X_j\right].$$

Proof From the definition of the variance, we can write $\operatorname{Var}[W_n] = E[(W_n - E[W_n])^2]$. For convenience, let μ_i denote $E[X_i]$. Since $W_n = \sum_{i=1}^{n} X_n$ and $E[W_n] = \sum_{i=1}^{n} \mu_i$, we can write

$$\operatorname{Var}[W_n] = E\left[\left(\sum_{i=1}^{n}(X_i - \mu_i)\right)^2\right] = E\left[\sum_{i=1}^{n}(X_i - \mu_i)\sum_{j=1}^{n}(X_j - \mu_j)\right] \tag{6.2}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{Cov}\left[X_i, X_j\right]. \tag{6.3}$$

In terms of the random vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_n \end{bmatrix}'$, we see that $\operatorname{Var}[W_n]$ is the sum of all the elements of the covariance matrix $\mathbf{C_X}$. Recognizing that $\operatorname{Cov}[X_i, X_i] = \operatorname{Var}[X]$ and $\operatorname{Cov}[X_i, X_j] = \operatorname{Cov}[X_j, X_i]$, we place the diagonal terms of $\mathbf{C_X}$ in one sum and the off-diagonal terms (which occur in pairs) in another sum to arrive at the formula in the theorem.

When $X_1, \ldots, X_n$ are uncorrelated, $\operatorname{Cov}[X_i, X_j] = 0$ for $i \neq j$ and the variance of the sum is the sum of the variances:

Theorem 6.3 *When $X_1, \ldots, X_n$ are uncorrelated,*

$$\operatorname{Var}[W_n] = \operatorname{Var}[X_1] + \cdots + \operatorname{Var}[X_n].$$

Example 6.1 $X_0, X_1, X_2, \ldots$ is a sequence of random variables with expected values $E[X_i] = 0$ and covariances, $\text{Cov}[X_i, X_j] = 0.8^{|i-j|}$. Find the expected value and variance of a random variable Y_i defined as the sum of three consecutive values of the random sequence

$$Y_i = X_i + X_{i-1} + X_{i-2}. \tag{6.4}$$

Theorem 6.1 implies that

$$E[Y_i] = E[X_i] + E[X_{i-1}] + E[X_{i-2}] = 0. \tag{6.5}$$

Applying Theorem 6.2, we obtain for each i,

$$\begin{aligned}\text{Var}[Y_i] &= \text{Var}[X_i] + \text{Var}[X_{i-1}] + \text{Var}[X_{i-2}] \\ &\quad + 2\,\text{Cov}[X_i, X_{i-1}] + 2\,\text{Cov}[X_i, X_{i-2}] + 2\,\text{Cov}[X_{i-1}, X_{i-2}].\end{aligned} \tag{6.6}$$

We next note that $\text{Var}[X_i] = \text{Cov}[X_i, X_i] = 0.8^{i-i} = 1$ and that

$$\text{Cov}[X_i, X_{i-1}] = \text{Cov}[X_{i-1}, X_{i-2}] = 0.8^1, \qquad \text{Cov}[X_i, X_{i-2}] = 0.8^2. \tag{6.7}$$

Therefore

$$\text{Var}[Y_i] = 3 \times 0.8^0 + 4 \times 0.8^1 + 2 \times 0.8^2 = 7.48. \tag{6.8}$$

Example 6.2 At a party of $n \geq 2$ people, each person throws a hat in a common box. The box is shaken and each person blindly draws a hat from the box without replacement. We say a match occurs if a person draws his own hat. What are the expected value and variance of V_n, the number of matches?

Let X_i denote an indicator random variable such that

$$X_i = \begin{cases} 1 & \text{person } i \text{ draws his hat,} \\ 0 & \text{otherwise.} \end{cases} \tag{6.9}$$

The number of matches is $V_n = X_1 + \cdots + X_n$. Note that the X_i are generally not independent. For example, with $n = 2$ people, if the first person draws his own hat, then the second person must also draw her own hat. Note that the ith person is equally likely to draw any of the n hats, thus $P_{X_i}(1) = 1/n$ and $E[X_i] = P_{X_i}(1) = 1/n$. Since the expected value of the sum always equals the sum of the expected values,

$$E[V_n] = E[X_1] + \cdots + E[X_n] = n(1/n) = 1. \tag{6.10}$$

To find the variance of V_n, we will use Theorem 6.2. The variance of X_i is

$$\mathrm{Var}[X_i] = E\left[X_i^2\right] - (E[X_i])^2 = \frac{1}{n} - \frac{1}{n^2}. \tag{6.11}$$

To find $\mathrm{Cov}[X_i, X_j]$, we observe that

$$\mathrm{Cov}[X_i, X_j] = E[X_i X_j] - E[X_i] E[X_j]. \tag{6.12}$$

Note that $X_i X_j = 1$ if and only if $X_i = 1$ and $X_j = 1$, and that $X_i X_j = 0$ otherwise. Thus

$$E[X_i X_j] = P_{X_i, X_j}(1, 1) = P_{X_i|X_j}(1|1) P_{X_j}(1). \tag{6.13}$$

Given $X_j = 1$, that is, the jth person drew his own hat, then $X_i = 1$ if and only if the ith person draws his own hat from the $n-1$ other hats. Hence $P_{X_i|X_j}(1|1) = 1/(n-1)$ and

$$E[X_i X_j] = \frac{1}{n(n-1)}, \qquad \mathrm{Cov}[X_i, X_j] = \frac{1}{n(n-1)} - \frac{1}{n^2}. \tag{6.14}$$

Finally, we can use Theorem 6.2 to calculate

$$\mathrm{Var}[V_n] = n\,\mathrm{Var}[X_i] + n(n-1)\,\mathrm{Cov}\left[X_i, X_j\right] = 1. \tag{6.15}$$

That is, both the expected value and variance of V_n are 1, no matter how large n is!

Example 6.3 Continuing Example 6.2, suppose each person immediately returns to the box the hat that he or she drew. What is the expected value and variance of V_n, the number of matches?

In this case the indicator random variables X_i are iid because each person draws from the same bin containing all n hats. The number of matches $V_n = X_1 + \cdots + X_n$ is the sum of n iid random variables. As before, the expected value of V_n is

$$E[V_n] = nE\left[X_i\right] = 1. \tag{6.16}$$

In this case, the variance of V_n equals the sum of the variances,

$$\mathrm{Var}[V_n] = n\,\mathrm{Var}[X_i] = n\left(\frac{1}{n} - \frac{1}{n^2}\right) = 1 - \frac{1}{n}. \tag{6.17}$$

deterministic sum: distribution

Example 6.6 J and K are independent random variables with probability mass functions

$$P_J(j) = \begin{cases} 0.2 & j = 1, \\ 0.6 & j = 2, \\ 0.2 & j = 3, \\ 0 & \text{otherwise,} \end{cases} \qquad P_K(k) = \begin{cases} 0.5 & k = -1, \\ 0.5 & k = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{6.40}$$

Find the MGF of $M = J + K$? What are $E[M^3]$ and $P_M(m)$?

J and K have have moment generating functions

$$\phi_J(s) = 0.2e^s + 0.6e^{2s} + 0.2e^{3s}, \qquad \phi_K(s) = 0.5e^{-s} + 0.5e^s. \tag{6.41}$$

Therefore, by Theorem 6.8, $M = J + K$ has MGF

$$\phi_M(s) = \phi_J(s)\phi_K(s) = 0.1 + 0.3e^s + 0.2e^{2s} + 0.3e^{3s} + 0.1e^{4s}. \tag{6.42}$$

To find the third moment of M, we differentiate $\phi_M(s)$ three times:

$$E\left[M^3\right] = \left.\frac{d^3\phi_M(s)}{ds^3}\right|_{s=0} \tag{6.43}$$

$$= 0.3e^s + 0.2(2^3)e^{2s} + 0.3(3^3)e^{3s} + 0.1(4^3)e^{4s}\Big|_{s=0} = 16.4. \tag{6.44}$$

The value of $P_M(m)$ at any value of m is the coefficient of e^{ms} in $\phi_M(s)$:

$$\phi_M(s) = E\left[e^{sM}\right] = \underbrace{0.1}_{P_M(0)} + \underbrace{0.3}_{P_M(1)} e^s + \underbrace{0.2}_{P_M(2)} e^{2s} + \underbrace{0.3}_{P_M(3)} e^{3s} + \underbrace{0.1}_{P_M(4)} e^{4s}. \tag{6.45}$$

The complete expression for the PMF of M is

$$P_M(m) = \begin{cases} 0.1 & m = 0, 4, \\ 0.3 & m = 1, 3, \\ 0.2 & m = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{6.46}$$

deterministic sum: distribution

Theorem 6.10 *The sum of n independent Gaussian random variables $W = X_1 + \cdots + X_n$ is a Gaussian random variable.*

Proof For convenience, let $\mu_i = E[X_i]$ and $\sigma_i^2 = \text{Var}[X_i]$. Since the X_i are independent, we know that

$$\phi_W(s) = \phi_{X_1}(s)\phi_{X_2}(s)\cdots\phi_{X_n}(s) \tag{6.49}$$

$$= e^{s\mu_1+\sigma_1^2 s^2/2} e^{s\mu_2+\sigma_2^2 s^2/2} \cdots e^{s\mu_n+\sigma_n^2 s^2/2} \tag{6.50}$$

$$= e^{s(\mu_1+\cdots+\mu_n)+(\sigma_1^2+\cdots+\sigma_n^2)s^2/2}. \tag{6.51}$$

From Equation (6.51), we observe that $\phi_W(s)$ is the moment generating function of a Gaussian random variable with expected value $\mu_1 + \cdots + \mu_n$ and variance $\sigma_1^2 + \cdots + \sigma_n^2$.

In general, the sum of independent random variables in one family is a different kind of random variable. The following theorem shows that the Erlang (n, λ) random variable is the sum of n independent exponential (λ) random variables.

Theorem 6.11 *If $X_1, \ldots, X_n$ are iid exponential (λ) random variables, then $W = X_1 + \cdots + X_n$ has the Erlang PDF*

$$f_W(w) = \begin{cases} \frac{\lambda^n w^{n-1} e^{-\lambda w}}{(n-1)!} & w \geq 0, \\ 0 & \textit{otherwise}. \end{cases}$$

Proof In Table 6.1 we observe that each X_i has MGF $\phi_X(s) = \lambda/(\lambda - s)$. By Theorem 6.8, W has MGF

$$\phi_W(s) = \left(\frac{\lambda}{\lambda - s}\right)^n. \tag{6.52}$$

Returning to Table 6.1, we see that W has the MGF of an Erlang (n, λ) random variable.

random sum: distribution

Example 6.7 At a bus terminal, count the number of people arriving on buses during one minute. If the number of people on the ith bus is K_i and the number of arriving buses is N, then the number of people arriving during the minute is

$$R = K_1 + \cdots + K_N. \tag{6.56}$$

In general, the number N of buses that arrive is a random variable. Therefore, R is a random sum of random variables.

In the preceding example we can use the methods of Chapter 4 to find the joint PMF $P_{N,R}(n,r)$. However, we are not able to find a simple closed form expression for the PMF

$P_R(r)$. On the other hand, we see in the next theorem that it is possible to express the probability model of R as a formula for the moment generating function $\phi_R(s)$.

random sum: distribution

Theorem 6.12 *Let $\{X_1, X_2, \ldots\}$ be a collection of iid random variables, each with MGF $\phi_X(s)$, and let N be a nonnegative integer-valued random variable that is independent of $\{X_1, X_2, \ldots\}$. The random sum $R = X_1 + \cdots + X_N$ has moment generating function*

$$\phi_R(s) = \phi_N(\ln \phi_X(s)).$$

Proof To find $\phi_R(s) = E[e^{sR}]$, we first find the conditional expected value $E[e^{sR}|N = n]$. Because this expected value is a function of n, it is a random variable. Theorem 4.26 states that $\phi_R(s)$ is the expected value, with respect to N, of $E[e^{sR}|N = n]$:

$$\phi_R(s) = \sum_{n=0}^{\infty} E\left[e^{sR}|N = n\right] P_N(n) = \sum_{n=0}^{\infty} E\left[e^{s(X_1+\cdots+X_N)}|N = n\right] P_N(n). \tag{6.58}$$

Because the X_i are independent of N,

$$E\left[e^{s(X_1+\cdots+X_N)}|N = n\right] = E\left[e^{s(X_1+\cdots+X_n)}\right] = E\left[e^{sW}\right] = \phi_W(s). \tag{6.59}$$

In Equation (6.58), $W = X_1 + \cdots + X_n$. From Theorem 6.8, we know that $\phi_W(s) = [\phi_X(s)]^n$, implying

$$\phi_R(s) = \sum_{n=0}^{\infty} [\phi_X(s)]^n P_N(n). \tag{6.60}$$

We observe that we can write $[\phi_X(s)]^n = [e^{\ln \phi_X(s)}]^n = e^{[\ln \phi_X(s)]n}$. This implies

$$\phi_N(s) = E\left[e^{sN}\right] = \sum_{n \in S_N} e^{sn} P_N(n) \quad (6.27)$$

$$\phi_R(s) = \sum_{n=0}^{\infty} e^{[\ln \phi_X(s)]n} P_N(n). \tag{6.61}$$

Recognizing that this sum has the same form as the sum in Equation (6.27), we infer that the sum is $\phi_N(s)$ evaluated at $s = \ln \phi_X(s)$. Therefore, $\phi_R(s) = \phi_N(\ln \phi_X(s))$.

In the following example, we find the MGF of a random sum and then transform it to the PMF.

Example 6.9 The number of pages N in a fax transmission has a geometric PMF with expected value $1/q = 4$. The number of bits K in a fax page also has a geometric distribution with expected value $1/p = 10^5$ bits, independent of the number of bits in any other page and independent of the number of pages. Find the MGF and the PMF of B, the total number of bits in a fax transmission.

When the ith page has K_i bits, the total number of bits is the random sum $B = K_1 + \cdots + K_N$. Thus $\phi_B(s) = \phi_N(\ln \phi_K(s))$. From Table 6.1,

$$\phi_N(s) = \frac{qe^s}{1-(1-q)e^s}, \qquad \phi_K(s) = \frac{pe^s}{1-(1-p)e^s}. \tag{6.62}$$

To calculate $\phi_B(s)$, we substitute $\ln \phi_K(s)$ for every occurrence of s in $\phi_N(s)$. Equivalently, we can substitute $\phi_K(s)$ for every occurrence of e^s in $\phi_N(s)$. yields

$$\phi_B(s) = \frac{q\left(\frac{pe^s}{1-(1-p)e^s}\right)}{1-(1-q)\left(\frac{pe^s}{1-(1-p)e^s}\right)} = \frac{pqe^s}{1-(1-pq)e^s}. \tag{6.63}$$

By comparing $\phi_K(s)$ and $\phi_B(s)$, we see that B has the MGF of a geometric ($pq = 2.5 \times 10^{-5}$) random variable with expected value $1/(pq) = 400{,}000$ bits. Therefore, B has the geometric PMF

$$P_B(b) = \begin{cases} pq(1-pq)^{b-1} & b = 1, 2, \ldots, \\ 0 & \text{otherwise}, \end{cases} \tag{6.64}$$

Using Theorem 6.12, we can take derivatives of $\phi_N(\ln \phi_X(s))$ to find simple expressions for the expected value and variance of R.

Theorem 6.13 *For the random sum of iid random variables $R = X_1 + \cdots + X_N$,*

$$E[R] = E[N]\,E[X], \qquad \text{Var}[R] = E[N]\,\text{Var}[X] + \text{Var}[N]\,(E[X])^2.$$

Proof By the chain rule for derivatives,

$$\phi_R'(s) = \phi_N'(\ln \phi_X(s)) \frac{\phi_X'(s)}{\phi_X(s)}. \tag{6.65}$$

Since $\phi_X(0) = 1$, $\phi_N'(0) = E[N]$, and $\phi_X'(0) = E[X]$, evaluating the equation at $s = 0$ yields

$$E[R] = \phi_R'(0) = \phi_N'(0) \frac{\phi_X'(0)}{\phi_X(0)} = E[N]\,E[X]. \tag{6.66}$$

For the second derivative of $\phi_X(s)$, we have

$$\phi_R''(s) = \phi_N''(\ln \phi_X(s)) \left(\frac{\phi_X'(s)}{\phi_X(s)}\right)^2 + \phi_N'(\ln \phi_X(s)) \frac{\phi_X(s)\phi_X''(s) - \left[\phi_X'(s)\right]^2}{[\phi_X(s)]^2}. \tag{6.67}$$

The value of this derivative at $s = 0$ is

$$E\left[R^2\right] = E\left[N^2\right]\mu_X^2 + E[N]\left(E\left[X^2\right] - \mu_X^2\right). \tag{6.68}$$

Subtracting $(E[R])^2 = (\mu_N \mu_X)^2$ from both sides of this equation completes the proof.

We observe that $\text{Var}[R]$ contains two terms: the first term, $\mu_N \text{Var}[X]$, results from the randomness of X, while the second term, $\text{Var}[N]\mu_X^2$, is a consequence of the randomness of N. To see this, consider these two cases.

- Suppose N is deterministic such that $N = n$ every time. In this case, $\mu_N = n$ and $\text{Var}[N] = 0$. The random sum R is an ordinary deterministic sum $R = X_1 + \cdots + X_n$ and $\text{Var}[R] = n \text{Var}[X]$.
- Suppose N is random, but each X_i is a deterministic constant x. In this instance, $\mu_X = x$ and $\text{Var}[X] = 0$. Moreover, the random sum becomes $R = Nx$ and $\text{Var}[R] = x^2 \text{Var}[N]$.

We emphasize that Theorems 6.12 and 6.13 require that N be independent of the random variables $X_1, X_2, \ldots$. That is, the number of terms in the random sum cannot depend on the actual values of the terms in the sum.