6. Mean, Variance, Moments and Characteristic Functions

For a r.v X, its p.d.f $f_X(x)$ represents complete information about it, and for any Borel set B on the x-axis

$$P\big(X(\xi) \in B\big) = \int_B f_X(x)\,dx. \tag{6-1}$$

Note that $f_X(x)$ represents very detailed information, and quite often it is desirable to characterize the r.v in terms of its average behavior. In this context, we will introduce two parameters - mean and variance - that are universally used to represent the overall properties of the r.v and its p.d.f.

Mean or the **Expected Value** of a r.v X is defined as

$$\eta_X = \overline{X} = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)dx. \tag{6-2}$$

If X is a discrete-type r.v, then using (3-25) we get

$$\eta_X = \overline{X} = E(X) = \int x \sum_i p_i \delta(x - x_i)dx = \sum_i x_i p_i \underbrace{\int \delta(x - x_i)dx}_{1}$$

$$= \sum_i x_i p_i = \sum_i x_i \, P(X = x_i). \tag{6-3}$$

Mean represents the average (mean) value of the r.v in a very large number of trials. For example if $X \sim U(a,b)$, then using (3-31) ,

$$E(X) = \int_a^b \frac{x}{b-a}dx = \frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \tag{6-4}$$

is the midpoint of the interval (a,b).

On the other hand if X is exponential with parameter λ as in (3-32), then

$$E(X) = \int_0^\infty \frac{x}{\lambda} e^{-x/\lambda} dx = \lambda \int_0^\infty y e^{-y} dy = \lambda, \tag{6-5}$$

implying that the parameter λ in (3-32) represents the mean value of the exponential r.v.

Similarly if X is Poisson with parameter λ as in (3-45), using (6-3), we get

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

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \tag{6-6}$$

Thus the parameter λ in (3-45) also represents the mean of the Poisson r.v.

In a similar manner, if X is binomial as in (3-44), then its mean is given by

$$E(X) = \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k \frac{n!}{(n-k)!\,k!} p^k q^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n!}{(n-k)!\,(k-1)!} p^k q^{n-k} = np \sum_{i=0}^{n-1} \frac{(n-1)!}{(n-i-1)!\,i!} p^i q^{n-i-1} = np(p+q)^{n-1} = np. \tag{6-7}$$

Thus np represents the mean of the binomial r.v in (3-44).

For the normal r.v in (3-29),

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (y+\mu) e^{-y^2/2\sigma^2} dy$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{\int_{-\infty}^{+\infty} y e^{-y^2/2\sigma^2} dy}_{0} + \mu \cdot \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-y^2/2\sigma^2} dy}_{1} = \mu. \tag{6-8}$$

Thus the first parameter in $X \sim N(\mu, \sigma^2)$ is infact the mean of the Gaussian r.v X. Given $X \sim f_X(x)$, suppose $Y = g(X)$ defines a new r.v with p.d.f $f_Y(y)$. Then from the previous discussion, the new r.v Y has a mean μ_Y given by (see (6-2))

$$\mu_Y = E(Y) = \int_{-\infty}^{+\infty} y \, f_Y(y) dy. \tag{6-9}$$

From (6-9), it appears that to determine $E(Y)$, we need to determine $f_Y(y)$. However this is not the case if only $E(Y)$ is the quantity of interest. Recall that for any y, $\Delta y > 0$

$$P\left(y < Y \le y + \Delta y\right) = \sum_i P\left(x_i < X \le x_i + \Delta x_i\right), \tag{6-10}$$

where x_i represent the multiple solutions of the equation $y = g(x_i)$. But(6-10) can be rewritten as

$$f_Y(y)\Delta y = \sum_i f_X(x_i)\Delta x_i, \tag{6-11}$$

where the $(x_i,\ x_i+\Delta x_i)$ terms form nonoverlapping intervals. Hence

$$y\ f_Y(y)\Delta y = \sum_i y\ f_X(x_i)\Delta x_i = \sum_i g(x_i) f_X(x_i)\Delta x_i, \qquad (6\text{-}12)$$

and hence as *Δy* covers the entire y-axis, the corresponding *Δx*'s are nonoverlapping, and they cover the entire *x*-axis. Hence, in the limit as $\Delta y \to 0$, integrating both sides of (6-12), we get the useful formula

$$E(Y) = E\big(g(X)\big) = \int_{-\infty}^{+\infty} y\ f_Y(y)dy = \int_{-\infty}^{+\infty} g(x) f_X(x)dx. \qquad (6\text{-}13)$$

In the discrete case, (6-13) reduces to

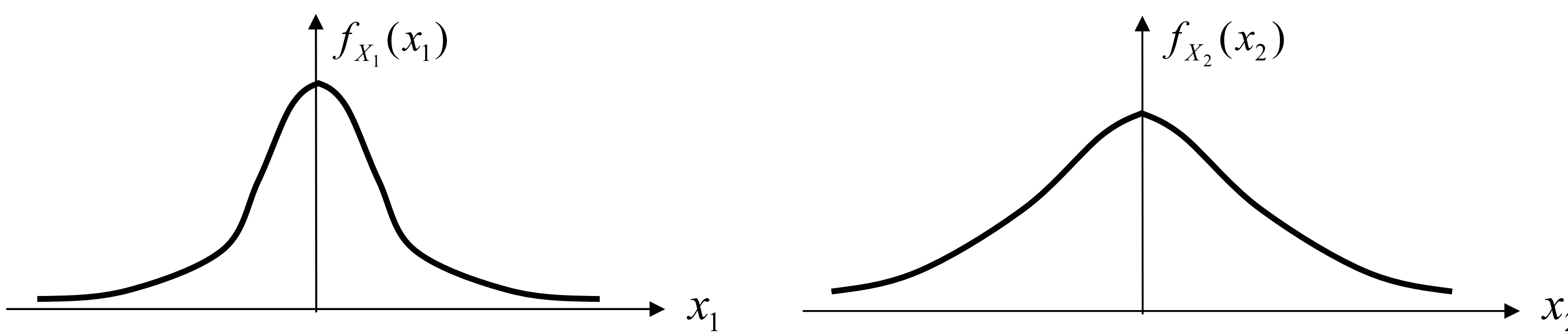
$$E(Y) = \sum_i g(x_i)P(X = x_i). \qquad (6\text{-}14)$$

From (6-13)-(6-14), $f_Y(y)$ is not required to evaluate $E(Y)$ for $Y = g(X)$. We can use (6-14) to determine the mean of $Y = X^2$, where *X* is a Poisson r.v. Using (3-45)

$$
\begin{aligned}
E\left(X^2\right) &= \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{(k-1)!} = e^{-\lambda}\sum_{i=0}^{\infty}(i+1)\frac{\lambda^{i+1}}{i!} \\
&= \lambda e^{-\lambda}\left(\sum_{i=0}^{\infty} i\frac{\lambda^i}{i!} + \sum_{i=0}^{\infty}\frac{\lambda^i}{i!}\right) = \lambda e^{-\lambda}\left(\sum_{i=1}^{\infty} i\frac{\lambda^i}{i!} + e^{\lambda}\right) \\
&= \lambda e^{-\lambda}\left(\sum_{i=1}^{\infty}\frac{\lambda^i}{(i-1)!} + e^{\lambda}\right) = \lambda e^{-\lambda}\left(\sum_{m=0}^{\infty}\frac{\lambda^{m+1}}{m!} + e^{\lambda}\right) \\
&= \lambda e^{-\lambda}\left(\lambda e^{\lambda} + e^{\lambda}\right) = \lambda^2 + \lambda. \qquad (6\text{-}15)
\end{aligned}
$$

In general, $E\left(X^k\right)$ is known as the *k*th moment of r.v X. Thus if $X \sim P(\lambda)$, its second moment is given by (6-15).

Mean alone will not be able to truly represent the p.d.f of any r.v. To illustrate this, consider the following scenario: Consider two Gaussian r.vs $X_1 \sim N(0,1)$ and $X_2 \sim N(0,10)$. Both of them have the same mean $\mu = 0.$ However, as Fig. 6.1 shows, their p.d.fs are quite different. One is more concentrated around the mean, whereas the other one (X_2) has a wider spread. Clearly, we need atleast an additional parameter to measure this spread around the mean!



(a) $\sigma^2 = 1$ (b) $\sigma^2 = 10$

Fig.6.1

For a r.v X with mean μ, $X-\mu$ represents the deviation of the r.v from its mean. Since this deviation can be either positive or negative, consider the quantity $(X-\mu)^2$, and its average value $E[(X-\mu)^2]$ represents the average mean square deviation of X around its mean. Define

$$\sigma_X^2 \stackrel{\Delta}{=} E[(X-\mu)^2] > 0. \tag{6-16}$$

With $g(X)=(X-\mu)^2$ and using (6-13) we get

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx > 0. \tag{6-17}$$

σ_X^2 is known as the variance of the r.v X, and its square root $\sigma_X = \sqrt{E(X-\mu)^2}$ is known as the standard deviation of X. Note that the standard deviation represents the root mean square spread of the r.v X around its mean μ.

Expanding (6-17) and using the linearity of the integrals, we get

$$Var(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} \left(x^2 - 2x\mu + \mu^2\right) f_X(x)\,dx$$

$$= \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx - 2\mu \int_{-\infty}^{+\infty} x\, f_X(x)\,dx + \mu^2$$

$$= E\left(X^2\right) - \mu^2 = E\left(X^2\right) - [E(X)]^2 = \overline{X^2} - \overline{X}^2. \quad (6\text{-}18)$$

Alternatively, we can use (6-18) to compute σ_X^2.

Thus , for example, returning back to the Poisson r.v in (3-45), using (6-6) and (6-15), we get

$$\sigma_X^2 = \overline{X^2} - \overline{X}^2 = \left(\lambda^2 + \lambda\right) - \lambda^2 = \lambda. \quad (6\text{-}19)$$

Thus for a Poisson r.v, mean and variance are both equal to its parameter λ.

To determine the variance of the normal r.v $N(\mu, \sigma^2)$, we can use (6-16). Thus from (3-29)

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx. \tag{6-20}$$

To simplify (6-20), we can make use of the identity

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

for a normal p.d.f. This gives

$$\int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}\,\sigma. \tag{6-21}$$

Differentiating both sides of (6-21) with respect to σ, we get

$$\int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma^3} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}$$

or

$$\int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \sigma^2, \tag{6-22}$$

which represents the $Var(X)$ in (6-20). Thus for a normal r.v as in (3-29)

$$Var(X) = \sigma^2 \tag{6-23}$$

and the second parameter in $N(\mu, \sigma^2)$ infact represents the variance of the Gaussian r.v. As Fig. 6.1 shows the larger the σ, the larger the spread of the p.d.f around its mean. Thus as the variance of a r.v tends to zero, it will begin to concentrate more and more around the mean ultimately behaving like a constant.

Moments: As remarked earlier, in general

$$m_n = \overline{X^n} = E(X^n), \quad n \geq 1 \tag{6-24}$$

are known as the moments of the r.v *X*, and

$$\mu_n = E[(X-\mu)^n] \tag{6-25}$$

are known as the central moments of X. Clearly, the mean $\mu = m_1$, and the variance $\sigma^2 = \mu_2$. It is easy to relate m_n and μ_n. Infact

$$\mu_n = E[(X-\mu)^n] = E\left(\sum_{k=0}^{n}\binom{n}{k}X^k(-\mu)^{n-k}\right)$$

$$= \sum_{k=0}^{n}\binom{n}{k}E\left(X^k\right)(-\mu)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}\ m_k\ (-\mu)^{n-k}. \tag{6-26}$$

In general, the quantities

$$E[(X-a)^n] \tag{6-27}$$

are known as the generalized moments of X about a, and

$$E[|X|^n] \tag{6-28}$$

are known as the absolute moments of X.

For example, if $X \sim N(0,\sigma^2)$, then it can be shown that

$$E(X^n) = \begin{cases} 0, & n \text{ odd}, \\ 1\cdot 3 \cdots (n-1)\sigma^n, & n \text{ even}. \end{cases} \tag{6-29}$$

$$E(|X|^n) = \begin{cases} 1\cdot 3 \cdots (n-1)\sigma^n, & n \text{ even}, \\ 2^k k! \sigma^{2k+1}\sqrt{2/\pi}, & n = (2k+1), \text{ odd}. \end{cases} \tag{6-30}$$

Direct use of (6-2), (6-13) or (6-14) is often a tedious procedure to compute the mean and variance, and in this context, the notion of the characteristic function can be quite helpful.

Characteristic Function

The characteristic function of a r.v X is defined as

$$\Phi_X(\omega) \overset{\Delta}{=} E\left(e^{jX\omega}\right) = \int_{-\infty}^{+\infty} e^{jx\omega} f_X(x)\,dx. \tag{6-31}$$

Thus $\Phi_X(0) = 1,$ and $|\Phi_X(\omega)| \le 1$ for all ω.

For discrete r.vs the characteristic function reduces to

$$\Phi_X(\omega) = \sum_k e^{jk\omega} P(X = k). \tag{6-32}$$

Thus for example, if $X \sim P(\lambda)$ as in (3-45), then its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} e^{jk\omega} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{j\omega})^k}{k!} = e^{-\lambda} e^{\lambda e^{j\omega}} = e^{\lambda(e^{j\omega}-1)}. \tag{6-33}$$

Similarly, if X is a binomial r.v as in (3-44), its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{n} e^{jk\omega} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^{j\omega})^k q^{n-k} = (pe^{j\omega} + q)^n. \tag{6-34}$$

To illustrate the usefulness of the characteristic function of a r.v in computing its moments, first it is necessary to derive the relationship between them. Towards this, from (6-31)

$$\Phi_X(\omega) = E\left(e^{jX\omega}\right) = E\left[\sum_{k=0}^{\infty} \frac{(j\omega X)^k}{k!}\right] = \sum_{k=0}^{\infty} j^k \frac{E(X^k)}{k!}\omega^k$$

$$= 1 + jE(X)\omega + j^2 \frac{E(X^2)}{2!}\omega^2 + \cdots + j^k \frac{E(X^k)}{k!}\omega^k + \cdots. \qquad (6\text{-}35)$$

Taking the first derivative of (6-35) with respect to ω, and letting it to be equal to zero, we get

$$\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0} = jE(X) \quad \text{or} \quad E(X) = \frac{1}{j}\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0}. \qquad (6\text{-}36)$$

Similarly, the second derivative of (6-35) gives

$$E(X^2) = \frac{1}{j^2}\left.\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2}\right|_{\omega=0}, \qquad (6\text{-}37)$$

and repeating this procedure k times, we obtain the kth moment of X to be

$$E(X^k) = \frac{1}{j^k} \frac{\partial^k \Phi_X(\omega)}{\partial \omega^k}\bigg|_{\omega=0}, \quad k \geq 1. \tag{6-38}$$

We can use (6-36)-(6-38) to compute the mean, variance and other higher order moments of any r.v X. For example, if $X \sim P(\lambda)$, then from (6-33)

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = e^{-\lambda} e^{\lambda e^{j\omega}} \lambda j e^{j\omega}, \tag{6-39}$$

so that from (6-36)

$$E(X) = \lambda, \tag{6-40}$$

which agrees with (6-6). Differentiating (6-39) one more time, we get

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = e^{-\lambda}\left(e^{\lambda e^{j\omega}}(\lambda j e^{j\omega})^2 + e^{\lambda e^{j\omega}} \lambda j^2 e^{j\omega}\right), \tag{6-41}$$

so that from (6-37)

$$E(X^2) = \lambda^2 + \lambda, \tag{6-42}$$

which again agrees with (6-15). Notice that compared to the tedious calculations in (6-6) and (6-15), the efforts involved in (6-39) and (6-41) are very minimal.

We can use the characteristic function of the binomial r.v $B(n, p)$ in (6-34) to obtain its variance. Direct differentiation of (6-34) gives

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = jnpe^{j\omega}(pe^{j\omega} + q)^{n-1} \tag{6-43}$$

so that from (6-36), $E(X) = np$ as in (6-7).

One more differentiation of (6-43) yields

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = j^2 np\left(e^{j\omega}(pe^{j\omega}+q)^{n-1} + (n-1)pe^{j2\omega}(pe^{j\omega}+q)^{n-2}\right) \qquad (6\text{-}44)$$

and using (6-37), we obtain the second moment of the binomial r.v to be

$$E(X^2) = np\left(1+(n-1)p\right) = n^2p^2 + npq. \qquad (6\text{-}45)$$

Together with (6-7), (6-18) and (6-45), we obtain the variance of the binomial r.v to be

$$\sigma_X^2 = E(X^2) - \left[E(X)\right]^2 = n^2p^2 + npq - n^2p^2 = npq. \qquad (6\text{-}46)$$

To obtain the characteristic function of the Gaussian r.v, we can make use of (6-31). Thus if $X \sim N(\mu, \sigma^2)$, then

$$\Phi_X(\omega) = \int_{-\infty}^{+\infty} e^{j\omega x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \quad (\text{Let } x - \mu = y)$$

$$= e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{j\omega y} e^{-y^2/2\sigma^2} dy = e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-y/2\sigma^2 (y - j2\sigma^2\omega)} dy$$

$$(\text{Let } y - j\sigma^2\omega = u \text{ so that } y = u + j\sigma^2\omega)$$

$$= e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-(u + j\sigma^2\omega)(u - j\sigma^2\omega)/2\sigma^2} du$$

$$= e^{j\mu\omega} e^{-\sigma^2\omega^2/2} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-u^2/2\sigma^2} du = e^{(j\mu\omega - \sigma^2\omega^2/2)}. \tag{6-47}$$

Notice that the characteristic function of a Gaussian r.v itself has the "Gaussian" bell shape. Thus if $X \sim N(0, \sigma^2)$, then

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}, \tag{6-48}$$

and

$$\Phi_X(\omega) = e^{-\sigma^2\omega^2/2}. \tag{6-49}$$

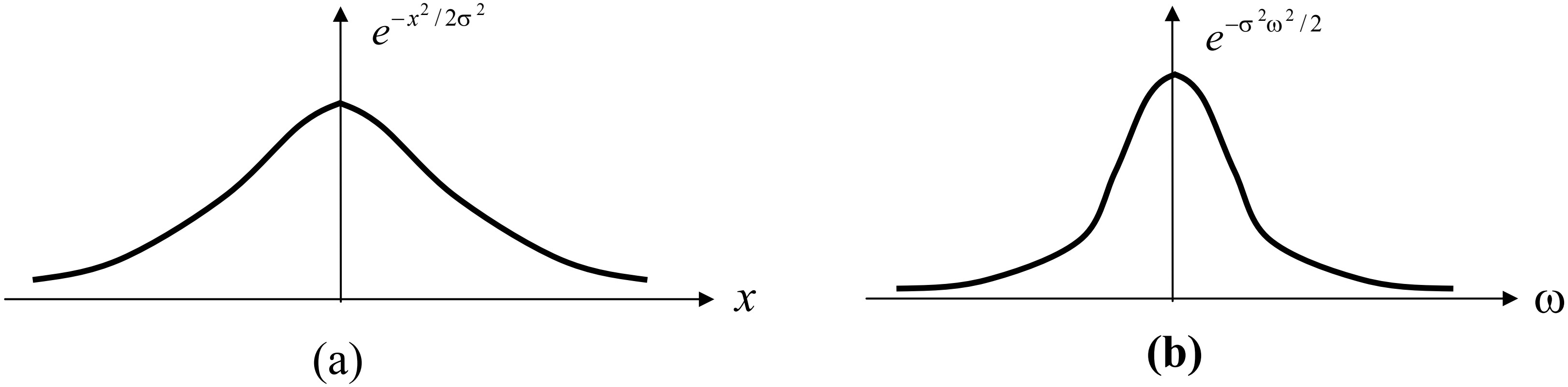


Fig. 6.2

From Fig. 6.2, the reverse roles of σ^2 in $f_X(x)$ and $\Phi_X(\omega)$ are noteworthy $(\sigma^2 \text{ vs } \frac{1}{\sigma^2})$.

In some cases, mean and variance may not exist. For example, consider the Cauchy r.v defined in (3-39). With

$$f_X(x) = \frac{(\alpha/\pi)}{\alpha^2 + x^2},$$

$$E(X^2) = \frac{\alpha}{\pi}\int_{-\infty}^{+\infty} \frac{x^2}{\alpha^2 + x^2}\,dx = \frac{\alpha}{\pi}\int_{-\infty}^{+\infty}\left(1 - \frac{\alpha^2}{\alpha^2 + x^2}\right)dx = \infty, \quad (6\text{-}50)$$

clearly diverges to infinity. Similarly

$$E(X) = \frac{\alpha}{\pi} \int_{-\infty}^{+\infty} \frac{x}{\alpha^2 + x^2} dx. \tag{6-51}$$

To compute (6-51), let us examine its one sided factor $\int_0^{+\infty} \frac{x}{\alpha^2 + x^2} dx.$ With $x = \alpha \tan\theta$

$$\int_0^{+\infty} \frac{x}{\alpha^2 + x^2} dx = \int_0^{\pi/2} \frac{\alpha \tan\theta}{\alpha^2 \sec^2\theta} \alpha \sec^2\theta \, d\theta = \int_0^{\pi/2} \frac{\sin\theta}{\cos\theta} d\theta$$

$$= -\int_0^{\pi/2} \frac{d(\cos\theta)}{\cos\theta} = -\log\cos\theta \Big|_0^{\pi/2} = -\log\cos\frac{\pi}{2} = \infty, \tag{6-52}$$

indicating that the double sided integral in (6-51) does not converge and is undefined. From (6-50)-(6-52), the mean and variance of a Cauchy r.v are undefined.

We conclude this section with a bound that estimates the dispersion of the r.v beyond a certain interval centered around its mean. Since σ^2 measures the dispersion of the r.v X around its mean μ, we expect this bound to depend on σ^2 as well.

Chebychev Inequality

Consider an interval of width 2ε symmetrically centered around its mean μ as in Fig. 6.3. What is the probability that X falls outside this interval? We need

$$P(|X - \mu| \geq \varepsilon)\ ? \qquad (6\text{-}53)$$

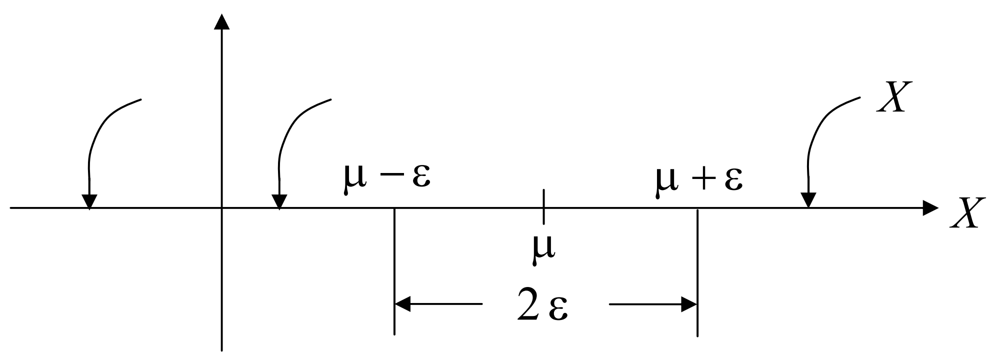


Fig. 6.3

To compute this probability, we can start with the definition of σ^2.

$$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx \geq \int_{|x-\mu|\geq\varepsilon} (x-\mu)^2 f_X(x)dx$$

$$\geq \int_{|x-\mu|\geq\varepsilon} \varepsilon^2 f_X(x)dx \geq \varepsilon^2 \int_{|x-\mu|\geq\varepsilon} f_X(x)dx \geq \varepsilon^2 P\left(|X-\mu| \geq \varepsilon\right). \quad \text{(6-54)}$$

From (6-54), we obtain the desired probability to be

$$P\left(|X-\mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}, \quad \text{(6-55)}$$

and (6-55) is known as the chebychev inequality. Interestingly, to compute the above probability bound the knowledge of $f_X(x)$ is not necessary. We only need σ^2, the variance of the r.v. In particular with $\varepsilon = k\sigma$ in (6-55) we obtain

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}. \quad \text{(6-56)}$$

Thus with $k = 3$, we get the probability of X being outside the 3σ interval around its mean to be 0.111 for any r.v. Obviously this cannot be a tight bound as it includes all r.vs. For example, in the case of a Gaussian r.v, from Table 4.1 $(\mu = 0, \sigma = 1)$

$$P(\,|\,X\,| \geq 3\sigma\,) = 0.0027\,. \tag{6-57}$$

which is much tighter than that given by (6-56). Chebychev inequality always underestimates the exact probability.

Moment Identities :

Suppose X is a discrete random variable that takes only nonnegative integer values. i.e.,

$$P(X = k) = p_k \geq 0, \quad k = 0, 1, 2, \cdots$$

Then

$$\sum_{k=0}^{\infty} P(X > k) = \sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} P(X = i) = \sum_{i=1}^{\infty} P(X = i) \sum_{k=0}^{i-1} 1$$

$$= \sum_{i=0}^{\infty} i\, P(X = i) = E(X) \qquad (6\text{-}58)$$

similarly

$$\sum_{k=0}^{\infty} k\, P(X > k) = \sum_{i=1}^{\infty} P(X = i) \sum_{k=0}^{i-1} k = \sum_{i=1}^{\infty} \frac{i(i-1)}{2} P(X = i) = \frac{E\{X(X-1)\}}{2}$$

which gives

$$E(X^2) = \sum_{i=1}^{\infty} i^2 P(X=i) = \sum_{k=0}^{\infty} (2k+1) P(X>k). \qquad (6\text{-}59)$$

Equations (6-58) – (6-59) are at times quite useful in simplifying calculations. For example, referring to the Birthday Pairing Problem [Example 2-20., Text], let X represent the minimum number of people in a group for a birthday pair to occur. The probability that "the first n people selected from that group have different birthdays" is given by [$P(B)$ in page 39, Text]

$$p_n = \prod_{k=1}^{n-1} (1 - \frac{k}{N}) \approx e^{-n(n-1)/2N}.$$

But the event the "the first n people selected have different birthdays" is the same as the event " $X > n$." Hence

$$P(X > n) \approx e^{-n(n-1)/2N}.$$

Using (6-58), this gives the mean value of X to be

$$\begin{aligned} E(X) &= \sum_{n=0}^{\infty} P(X > n) \approx \sum_{n=0}^{\infty} e^{-n(n-1)/2N} \approx \int_{-1/2}^{\infty} e^{-(x^2-1/4)/2N} dx \\ &= e^{(1/8N)} \int_{-1/2}^{\infty} e^{-x^2/2N} dx = e^{(1/8N)} \left\{ \frac{1}{2}\sqrt{2\pi N} + \int_0^{1/2} e^{-x^2/2N} dx \right\} \\ &\approx \sqrt{\pi N/2} + \frac{1}{2} = 24.44. \end{aligned} \tag{6-60}$$

Similarly using (6-59) we get

$$
\begin{aligned}
E(X^2) &= \sum_{n=0}^{\infty}(2n+1)P(X>n) \\
&= \sum_{n=0}^{\infty}(2n+1)e^{-n(n-1)/2N} = \int_{-1/2}^{\infty} 2\,(x+1)e^{-(x^2-1/4)/2N}dx \\
&= 2e^{(1/8N)}\left\{\int_0^{\infty} xe^{-x^2/2N}dx + \int_0^{1/2} xe^{-x^2/2N}dx\right\} + 2\int_{-1/2}^{\infty} e^{-(x^2-1/4)/2N}dx \\
&= 2\left\{\frac{\sqrt{2\pi\, N}}{2}\sqrt{\frac{2}{\pi}}\sqrt{N} + \frac{1}{8}\right\} + 2E(X) \\
&= 2N + \frac{1}{4} + \sqrt{2\pi\, N} + 1 = 2N + \sqrt{2\pi\, N} + \frac{5}{4} \\
&= 779.139.
\end{aligned}
$$

Thus

$$
Var(X) = E(X^2) - (E(X))^2 = 181.82
$$

which gives

$$\sigma_X \approx 13.48.$$

Since the standard deviation is quite high compared to the mean value, the actual number of people required for a birthday coincidence could be anywhere from 25 to 40.

Identities similar to (6-58)-(6-59) can be derived in the case of continuous random variables as well. For example, if X is a nonnegative random variable with density function $f_X(x)$ and distribution function $F_X(X)$, then

$$E\{X\} = \int_0^\infty x f_X(x)dx = \int_0^\infty \left(\int_0^x dy\right) f_X(x)dx$$

$$= \int_0^\infty \left(\int_y^\infty f_X(x)dx\right)dy = \int_0^\infty P(X > y)dy = \int_0^\infty P(X > x)dx$$

$$= \int_0^\infty \{1 - F_X(x)\}dx = \int_0^\infty R(x)dx, \qquad (6\text{-}61)$$

where

$$R(x) = 1 - F_X(x) \geq 0, \qquad x > 0.$$

Similarly

$$E\{X^2\} = \int_0^\infty x^2 f_X(x)dx = \int_0^\infty \left(\int_0^x 2y dy \right) f_X(x)dx$$

$$= 2\int_0^\infty \left(\int_y^\infty f_X(x)dx \right) y dy$$

$$= 2\int_0^\infty x R(x)dx.$$

A Baseball Trivia (Pete Rose and Dimaggio):

In 1978 Pete Rose set a national league record by hitting a string of 44 games during a 162 game baseball season. How unusual was that event?
As we shall see, that indeed was a rare event. In that context, we will answer the following question: What is the probability that someone in major league baseball will repeat that performance and possibly set a new record in the next 50 year period? The answer will put Pete Rose's accomplishment in the proper perspective.

Solution: As example 5-32 (Text) shows consecutive successes in n trials correspond to a run of length r in n

trials. From (5-133)-(5-134) text, we get the probability of r successive hits in n games to be

$$p_n = 1 - \alpha_{n,r} + p^r \alpha_{n-r,r} \tag{6-62}$$

where

$$\alpha_{n,r} = \sum_{k=0}^{\lfloor n/(r+1) \rfloor} \binom{n-kr}{k} (-1)^k (qp^r)^k \tag{6-63}$$

and p represents the probability of a hit in a game. Pete Rose's batting average is 0.303, and on the average since a batter shows up about four times/game, we get

$$\begin{aligned} p &= P(\text{at least one hit / game}) \\ &= 1 - \mathrm{P}(\text{no hit / game}) \\ &= 1 - (1 - 0.303)^4 = 0.76399 \end{aligned} \tag{6-64}$$

Substituting this value for p into the expressions (6-62)-(6-63) with $r = 44$ and $n = 162$, we can compute the desired probability $p_n.$ However since n is quite large compared to r, the above formula is hopelessly time consuming in its implementation, and it is preferable to obtain a good approximation for $p_n.$

Towards this, notice that the corresponding moment generating function $\phi(z)$ for $q_n = 1 - p_n$ in Eq. (5-130) Text, is rational and hence it can be expanded in partial fraction as

$$\phi(z) = \frac{1 - p^r z^r}{1 - z + qp^r z^{r+1}} = \sum_{k=1}^{r} \frac{a_k}{z - z_k}, \qquad (6\text{-}65)$$

where only r roots (out of $r+1$) are accounted for, since the root $z = 1/p$ is common to both the numerator and the denominator of $\phi(z).$ Here

$$a_k = \lim_{z \to z_k} \frac{(1 - p^r z^r)(z - z_k)}{1 - z + qp^r z^{r+1}}$$

$$= \lim_{z \to z_k} \frac{(1 - p^r z^r) - rp^r z^{r-1}(z - z_k)}{-1 + (r+1)qp^r z^r}$$

or

$$a_k = \frac{p^r z_k{}^r - 1}{1 - (r+1)qp^r z_k^r}, \qquad k = 1, 2, \cdots, r \tag{6-66}$$

From (6-65) – (6-66)

$$\phi(z) = \sum_{k=1}^{r} \frac{a_k}{(-z_k)} \, \frac{1}{1 - z / z_k} = \sum_{n=0}^{\infty} \underbrace{(\sum_{k=1}^{r} A_k z_k^{-(n+1)})}_{q_n} z^n \ \stackrel{\Delta}{=} \ \sum_{n=0}^{\infty} q_n z^n \tag{6-67}$$

where

$$A_k = -a_k = \frac{1 - p^r z_k^r}{1 - (r+1)qp^r z_k^r}$$

and

$$q_n = 1 - p_n = \sum_{k=1}^{r} A_k \; z_k^{-(n+1)}. \qquad (6\text{-}68)$$

However (fortunately), the roots $z_k, \;\; k = 1, 2, \cdots, r$ in (6-65)-(6-67) are all not of the same importance (in terms of their relative magnitude with respect to unity). Notice that since for large *n*, $z_k^{-(n+1)} \to 0$ for $|z_k| > 1,$ only the roots nearest to unity contribute to (6-68) as n becomes larger.

To examine the nature of the roots of the denominator

$$A(z) = z - 1 - qp^r z^{r+1}$$

in (6-65), note that (refer to Fig 6.1) $A(0) = -1 < 0,$ $A(1) = -qp^r > A(0),$ $A(1/p) = 0,$ $A(\infty) < 0$ implying that for $z \geq 0,$ $A(z)$ increases from –1 and reaches a positive maximum at z_0 given by

$$\left.\frac{dA(z)}{dz}\right|_{z-z_0} = 1 - qp^r(r+1)z_0^r = 0,$$

which gives

$$z_0^r = \frac{1}{qp^r(r+1)}. \tag{6-69}$$

There onwards $A(z)$ decreases to $-\infty$. Thus there are two positive roots for the equation $A(z)=0$ given by $z_1 < z_0$ and $z_2 = 1/p > 1$. Since $A(1) = -qp^r \approx 0$ but negative, by continuity z_1 has the form $z_1 = 1+\varepsilon,\ \ \varepsilon > 0$. (see Fig 6.1)

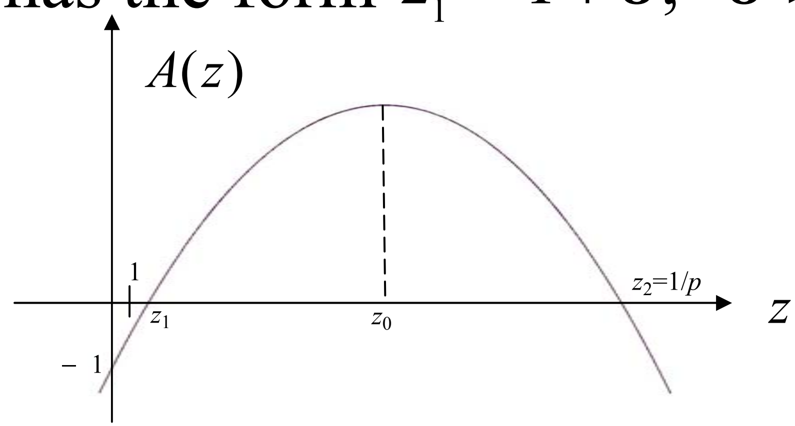


Fig 6.1 $A(z)$ for r odd

It is possible to obtain a bound for z_0 in (6-69). When P varies from 0 to 1, the maximum of $qp^r = (1-p)p^r$ is attained for $p = r/(r+1)$ and it equals $r^r/(r+1)^{r+1}$. Thus

$$qp^r \leq \frac{r^r}{(r+1)^{r+1}} \tag{6-70}$$

and hence substituting this into (6-69), we get

$$z_0 \geq \frac{r+1}{r} = 1 + \frac{1}{r}. \tag{6-71}$$

Hence it follows that the two positive roots of $A(z)$ satisfy

$$1 < z_1 < 1 + \frac{1}{r} < z_2 = \frac{1}{p} > 1. \tag{6-72}$$

Clearly, the remaining roots of $A(z)$ are complex if r is odd , and there is one negative root $-\alpha$ if r is even (see Fig 6.2). It is easy to show that the absolute value of *every* such complex or negative root is greater than $1/p > 1$.

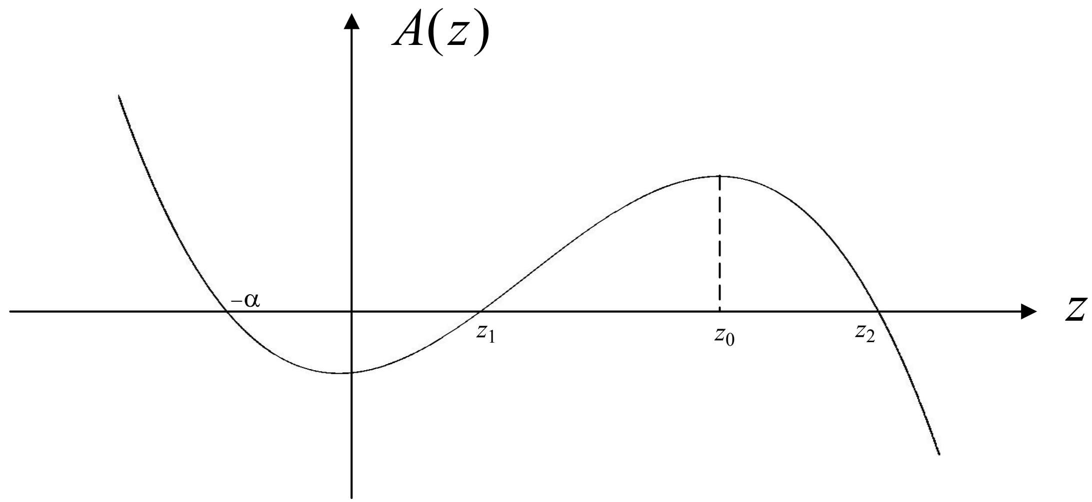


Fig 6.2 $A(z)$ for r even

To show this when r is even, suppose $-\alpha$ represents the negative root. Then

$$A(-\alpha) = -(\alpha + 1 - qp^r \alpha^{r+1}) = 0$$

so that the function

$$B(x) = x + 1 - qp^r x^{r+1} = A(x) + 2 \qquad (6\text{-}73)$$

starts positive, for $x > 0$ and increases till it reaches once again maximum at $z_0 \geq 1 + 1/r$ and then decreases to $-\infty$ through the root $x = \alpha > z_0 > 1$. Since $B(1/p) = 2$, we get $\alpha > 1/p > 1$, which proves our claim.

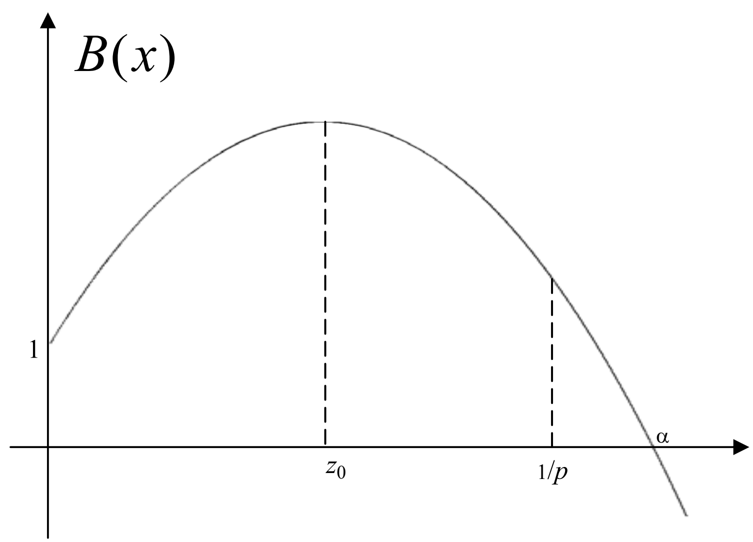


Fig 6.3 Negative root $B(\alpha) = 0$

Finally if $z = \rho\, e^{j\theta}$ is a complex root of $A(z)$, then

$$A(\rho\, e^{j\theta}) = \rho\, e^{j\theta} - 1 - qp^r \rho^{r+1} e^{j(r+1)\theta} = 0 \tag{6-74}$$

so that

$$\rho = | 1 + qp^r \rho^{r+1} e^{j(r+1)\theta} | \le 1 + qp^r \rho^{r+1}$$

or

$$A(\rho) = \rho - 1 - qp^r \rho^{r+1} < 0.$$

Thus from (6-72), ρ belongs to either the interval $(0,\ z_1)$ or the interval $(\frac{1}{p}, \infty)$ in Fig 6.1. Moreover , by equating the imaginary parts in (6-74) we get

$$qp^r \rho^r \frac{\sin(r+1)\theta}{\sin\theta} = 1. \tag{6-75}$$

But

$$\left|\frac{\sin(r+1)\theta}{\sin\theta}\right| \le r+1, \qquad (6\text{-}76)$$

equality being excluded if $\theta \neq 0$. Hence from (6-75)-(6-76) and (6-70)

$$(r+1)qp^r \rho^r > 1 \;\Rightarrow\; \rho^r > \frac{1}{(r+1)qp^r} = z_0^r > \left(\frac{r+1}{r}\right)^r$$

or

$$\rho > z_0 \ge 1 + \frac{1}{r}.$$

But $z_1 < z_0$. As a result ρ lies in the interval $(\frac{1}{p}, \infty)$ only. Thus

$$\rho > \frac{1}{p} > 1. \qquad (6\text{-}77)$$

To summarize the two real roots of the polynomial $A(z)$ are given by

$$z_1 = 1 + \varepsilon, \quad \varepsilon > 0; \quad z_2 = \frac{1}{p} > 1, \tag{6-78}$$

and all other roots are (negative or complex) of the form

$$z_k = \rho\, e^{j\theta} \quad \text{where } \rho > \frac{1}{p} > 1. \tag{6-79}$$

Hence except for the first root z_1 (which is very close to unity), for all other roots

$$z_k^{-(n+1)} \quad \rightarrow \quad 0 \quad \text{rapidly for all } k.$$

As a result, the most dominant term in (6-68) is the first term, and the contributions from all other terms to q_n in (6-68) can be bounded by

$$\left|\sum_{k=2}^{r} A_k z_k^{-(n+1)}\right| \le \sum_{k=2}^{r} |A_k||z_k|^{-(n+1)}$$

$$\le \sum_{k=2}^{r} \frac{1-(p|z_k|)^r}{1-(r+1)q(p|z_k|)^r}\, p^{n+1}$$

$$\le \sum_{k=2}^{r} \frac{(p|z_k|)^r}{(r+1)q(p|z_k|)^r}\, p^{n+1}$$

$$= \frac{r-1}{r+1}\frac{p^{n+1}}{q} \le \frac{p^{n+1}}{q} \;\; \to \;\; 0. \tag{6-80}$$

Thus from (6-68), to an excellent approximation

$$q_n = A_1 z_1^{-(n+1)}. \tag{6-81}$$

This gives the desired probability to be

$$p_n = 1 - q_n = 1 - \left(\frac{1-(pz_1)^r}{1-(r+1)q(pz_1)^r} \right) z_1^{-(n+1)}. \qquad (6\text{-}82)$$

Notice that since the dominant root z_1 is very close to unity, an excellent closed form approximation for z_1 can be obtained by considering the first order Taylor series expansion for $A(z)$. In the immediate neighborhood of $z = 1$ we get

$$A(1+\varepsilon) = A(1) + A'(1)\varepsilon = -qp^r + (1-(r+1)qp^r)\varepsilon$$

so that $A(z_1) = A(1+\varepsilon) = 0$ gives

$$\varepsilon = \frac{qp^r}{1-(r+1)qp^r},$$

or

$$z_1 \approx 1 + \frac{qp^r}{1-(r+1)qp^r}. \qquad (6\text{-}83)$$

Returning back to Pete Rose's case, $p = 0.763989$, $r = 44$ gives the smallest positive root of the denominator polynomial

$$A(z) = z - 1 - qp^{44} z^{45}$$

to be

$$z_1 = 1.00000169360549.$$

(The approximation (6-83) gives $z_1 = 1.00000169360548$).
Thus with n = 162 in (6-82) we get

$$p_{162} = 0.0002069970 \qquad (6\text{-}84)$$

to be the probability for scoring 44 or more consecutive hits in 162 games for a player of Pete Rose's caliber – a very small probability indeed! In that sense it is a very rare event.

Assuming that during any baseball season there are on the average about $2 \times 25 = 50$ (?) such players over all major league baseball teams, we obtain [use Lecture #2, Eqs.(2-3)-(2-6) for the independence of 50 players]

$$P_1 = 1 - (1 - p_{162})^{50} = 0.0102975349$$

to be the probability that one of those players will hit the desired event. If we consider a period of 50 years, then the probability of *some* player hitting 44 or more consecutive games during one of these game seasons turns out to be

$$1 - (1 - P_1)^{50} = 0.40401874. \qquad (6\text{-}85)$$

(We have once again used the independence of the 50 seasons.)

Thus Pete Rose's 44 hit performance has a 60-40 chance of survival for about 50 years.From (6-85), rare events do indeed occur. In other words, *some* unlikely event is likely to happen.
However, as (6-84) shows a *particular* unlikely event – such as Pete Rose hitting 44 games in a sequence – is indeed rare.

Table 6.1 lists p_{162} for various values of r. From there, every reasonable batter should be able to hit at least 10 to 12 consecutive games during every season!

r	$p_n \quad ; \ n = 162$
44	0.000207
25	0.03928
20	0.14937
15	0.48933
10	0.95257

Table 6.1 Probability of r runs in n trials for p=0.76399.

As baseball fans well know, Dimaggio holds the record of consecutive game hitting streak at 56 games (1941). With a lifetime batting average of 0.325 for Dimaggio, the above calculations yield [use (6-64), (6-82)-(6-83)] the probability for that event to be

$$p_n = 0.0000504532. \tag{6-86}$$

Even over a 100 year period, with an average of 50 excellent hitters / season, the probability is only

$$1-(1-P_0)^{100} = 0.2229669 \tag{6-87}$$

(where $P_0 = 1-(1-p_n)^{50} = 0.00251954$) that someone will repeat or outdo Dimaggio's performance.Remember, 60 years have already passed by, and no one has done it yet!