4. Binomial Random Variable Approximations, Conditional Probability Density Functions and Stirling's Formula

Let X represent a Binomial r.v as in (3-42). Then from (2-30)

$$P\left(k_1 \le X \le k_2\right) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \tag{4-1}$$

Since the binomial coefficient $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$ grows quite rapidly with n, it is difficult to compute (4-1) for large n. In this context, two approximations are extremely useful.

4.1 **The Normal Approximation** (Demoivre-Laplace Theorem) Suppose $n \to \infty$ with p held fixed. Then for k in the $\sqrt{npq}$ neighborhood of np, we can approximate

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq}. \qquad (4\text{-}2)$$

Thus if k_1 and k_2 in (4-1) are within or around the neighborhood of the interval $(np - \sqrt{npq},\ np + \sqrt{npq})$, we can approximate the summation in (4-1) by an integration. In that case (4-1) reduces to

$$P(k_1 \le X \le k_2) = \int_{k_1}^{k_2} \frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq} dx = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \qquad (4\text{-}3)$$

where

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

We can express (4-3) in terms of the normalized integral

$$erf(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy = erf(-x) \qquad (4\text{-}4)$$

that has been tabulated extensively (See Table 4.1).

For example, if x_1 and x_2 are both positive ,we obtain

$$P\left(k_1 \leq X \leq k_2\right) = erf\,(x_2) - erf\,(x_1). \tag{4-5}$$

Example 4.1: A fair coin is tossed 5,000 times. Find the probability that the number of heads is between 2,475 to 2,525.

Solution: We need $P(2{,}475 \leq X \leq 2{,}525)$. Here *n* is large so that we can use the normal approximation. In this case $p = \frac{1}{2}$, so that $np = 2{,}500$ and $\sqrt{npq} \approx 35$. Since $np - \sqrt{npq} = 2{,}465$, and $np + \sqrt{npq} = 2{,}535$, the approximation is valid for $k_1 = 2{,}475$ and $k_2 = 2{,}525$. Thus

$$P\left(k_1 \leq X \leq k_2\right) = \int_{x_1}^{x_2} \frac{1}{2\pi} e^{-y^2/2} dy.$$

Here $x_1 = \dfrac{k_1 - np}{\sqrt{npq}} = -\dfrac{5}{7}, \quad x_2 = \dfrac{k_2 - np}{\sqrt{npq}} = \dfrac{5}{7}.$

$$\operatorname{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy = G(x) - \frac{1}{2}$$

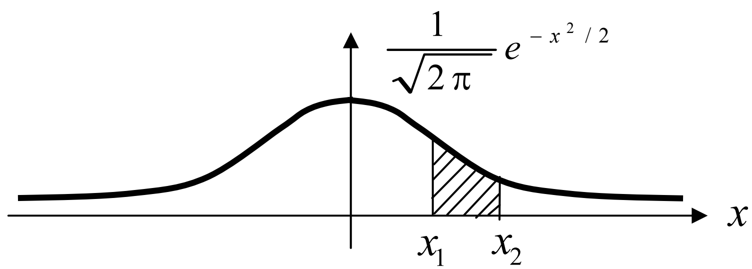
x	erf(x)	x	erf(x)	x	erf(x)	x	erf(x)
0.05	0.01994	0.80	0.28814	1.55	0.43943	2.30	0.48928
0.10	0.03983	0.85	0.30234	1.60	0.44520	2.35	0.49061
0.15	0.05962	0.90	0.31594	1.65	0.45053	2.40	0.49180
0.20	0.07926	0.95	0.32894	1.70	0.45543	2.45	0.49286
0.25	0.09871	1.00	0.34134	1.75	0.45994	2.50	0.49379
0.30	0.11791	1.05	0.35314	1.80	0.46407	2.55	0.49461
0.35	0.13683	1.10	0.36433	1.85	0.46784	2.60	0.49534
0.40	0.15542	1.15	0.37493	1.90	0.47128	2.65	0.49597
0.45	0.17364	1.20	0.38493	1.95	0.47441	2.70	0.49653
0.50	0.19146	1.25	0.39435	2.00	0.47726	2.75	0.49702
0.55	0.20884	1.30	0.40320	2.05	0.47982	2.80	0.49744
0.60	0.22575	1.35	0.41149	2.10	0.48214	2.85	0.49781
0.65	0.24215	1.40	0.41924	2.15	0.48422	2.90	0.49813
0.70	0.25804	1.45	0.42647	2.20	0.48610	2.95	0.49841
0.75	0.27337	1.50	0.43319	2.25	0.48778	3.00	0.49865

Table 4.1

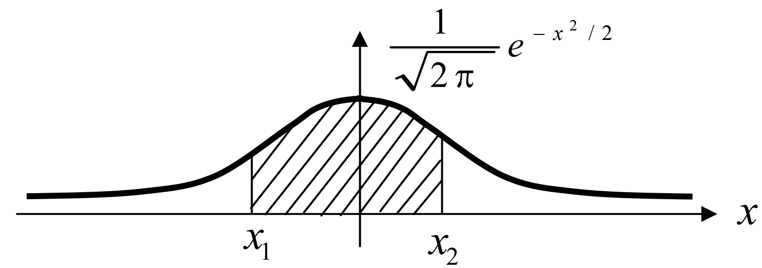
Since $x_1 < 0$, from Fig. 4.1(b), the above probability is given by $P(2{,}475 \le X \le 2{,}525) = \text{erf}(x_2) - \text{erf}(x_1) = \text{erf}(x_2) + \text{erf}(|x_1|)$

$$= 2\,\text{erf}\left(\frac{5}{7}\right) = 0.516,$$

where we have used Table 4.1 $(\text{erf}(0.7) = 0.258)$.



(a) $x_1 > 0,\ x_2 > 0$



(b) $x_1 < 0,\ x_2 > 0$

Fig. 4.1

4.2. **The Poisson Approximation**

As we have mentioned earlier, for large n, the Gaussian approximation of a binomial r.v is valid only if p is fixed, i.e., only if $np >> 1$ and $npq >> 1$. what if np is small, or if it does not increase with n?

Obviously that is the case if, for example, $p \to 0$ as $n \to \infty$, such that $np = \lambda$ is a fixed number.

Many random phenomena in nature in fact follow this pattern. Total number of calls on a telephone line, claims in an insurance company etc. tend to follow this type of behavior. Consider random arrivals such as telephone calls over a line. Let n represent the total number of calls in the interval $(0, T)$. From our experience, as $T \to \infty$ we have $n \to \infty$ so that we may assume $n = \mu T$. Consider a small interval of duration Δ as in Fig. 4.2. If there is only a single call coming in, the probability p of that single call occurring in that interval must depend on its relative size with respect to T.

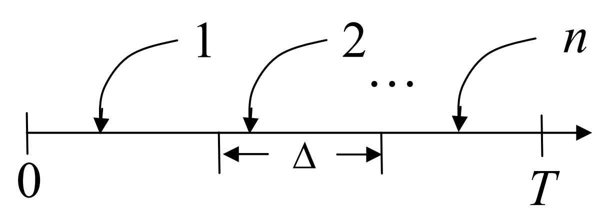


Fig. 4.2

Hence we may assume $p = \frac{\Delta}{T}.$ Note that $p \rightarrow 0$ as $T \rightarrow \infty$. However in this case $np = \mu T \cdot \frac{\Delta}{T} = \mu \Delta = \lambda$ is a constant, and the normal approximation is invalid here.

Suppose the interval Δ in Fig. 4.2 is of interest to us. A call inside that interval is a "success" (H), whereas one outside is a "failure" (T). This is equivalent to the coin tossing situation, and hence the probability $P_n(k)$ of obtaining k calls (in any order) in an interval of duration Δ is given by the binomial p.m.f. Thus

$$P_n(k) = \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k}, \tag{4-6}$$

and here as $n \rightarrow \infty$, $p \rightarrow 0$ such that $np = \lambda$. It is easy to obtain an excellent approximation to (4-6) in that situation. To see this, rewrite (4-6) as

$$P_n(k) = \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{(np)^k}{k!}(1-np/n)^{n-k}$$
$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{\lambda^k}{k!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}. \tag{4-7}$$

Thus

$$\lim_{n\to\infty,\, p\to 0,\, np=\lambda} P_n(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \tag{4-8}$$

since the finite products $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$ as well as $\left(1-\frac{\lambda}{n}\right)^k$ tend to unity as $n \to \infty$, and

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

The right side of (4-8) represents the Poisson p.m.f and the Poisson approximation to the binomial r.v is valid in situations where the binomial r.v parameters n and p diverge to two extremes $(n \to \infty, p \to 0)$ such that their product np is a constant.

Example 4.2: Winning a Lottery: Suppose two million lottery tickets are issued with 100 winning tickets among them. (a) If a person purchases 100 tickets, what is the probability of winning? (b) How many tickets should one buy to be 95% confident of having a winning ticket?

Solution: The probability of buying a winning ticket

$$p = \frac{\text{No. of winning tickets}}{\text{Total no. of tickets}} = \frac{100}{2\times 10^6} = 5\times 10^{-5}.$$

Here $n = 100$, and the number of winning tickets X in the n purchased tickets has an approximate Poisson distribution with parameter $\lambda = np = 100 \times 5 \times 10^{-5} = 0.005$. Thus

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},$$

and (a) Probability of winning $= P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda} \approx 0.005.$

(b) In this case we need $P(X \geq 1) \geq 0.95$.

$$P(X \geq 1) = 1 - e^{-\lambda} \geq 0.95 \quad \text{implies} \quad \lambda \geq \ln 20 = 3.$$

But $\lambda = np = n \times 5 \times 10^{-5} \geq 3$ or $n \geq 60{,}000$. Thus one needs to buy about 60,000 tickets to be 95% confident of having a winning ticket!

Example 4.3: A space craft has 100,000 components $(n \to \infty)$ The probability of any one component being defective is 2×10^{-5} $(p \to 0)$. The mission will be in danger if five or more components become defective. Find the probability of such an event.

Solution: Here n is large and p is small, and hence Poisson approximation is valid. Thus $np = \lambda = 100{,}000 \times 2 \times 10^{-5} = 2$, and the desired probability is given by

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - e^{-2} \sum_{k=0}^{4} \frac{\lambda^k}{k!}$$

$$= 1 - e^{-2}\left(1 + 2 + 2 + \frac{4}{3} + \frac{2}{3}\right) = 0.052.$$

Conditional Probability Density Function

For any two events A and B, we have defined the conditional probability of A given B as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0. \tag{4-9}$$

Noting that the probability distribution function $F_X(x)$ is given by

$$F_X(x) = P\{ X(\xi) \leq x \}, \tag{4-10}$$

we may define the conditional distribution of the r.v X given the event B as

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap B\}}{P(B)}. \tag{4-11}$$

Thus the definition of the conditional distribution depends on conditional probability, and since it obeys all probability axioms, it follows that the conditional distribution has the same properties as any distribution function. In particular

$$\begin{aligned} F_X(+\infty \mid B) &= \frac{P\{(X(\xi) \le +\infty) \cap B\}}{P(B)} = \frac{P(B)}{P(B)} = 1, \\ F_X(-\infty \mid B) &= \frac{P\{(X(\xi) \le -\infty) \cap B\}}{P(B)} = \frac{P(\phi)}{P(B)} = 0. \end{aligned} \tag{4-12}$$

Further

$$\begin{aligned} P(x_1 < X(\xi) \le x_2 \mid B) &= \frac{P\{(x_1 < X(\xi) \le x_2) \cap B\}}{P(B)} \\ &= F_X(x_2 \mid B) - F_X(x_1 \mid B), \end{aligned} \tag{4-13}$$

Since for $x_2 \geq x_1,$

$$\big(X(\xi) \leq x_2\big) = \big(X(\xi) \leq x_1\big) \cup \big(x_1 < X(\xi) \leq x_2\big). \qquad (4\text{-}14)$$

The conditional density function is the derivative of the conditional distribution function. Thus

$$f_X(x \mid B) = \frac{dF_X(x \mid B)}{dx}, \qquad (4\text{-}15)$$

and proceeding as in (3-26) we obtain

$$F_X(x \mid B) = \int_{-\infty}^{x} f_X(u \mid B)\,du. \qquad (4\text{-}16)$$

Using (4-16), we can also rewrite (4-13) as

$$P\big(x_1 < X(\xi) \leq x_2 \mid B\big) = \int_{x_1}^{x_2} f_X(x \mid B)\,dx. \qquad (4\text{-}17)$$

Example 4.4: Refer to example 3.2. Toss a coin and $X(T)=0$, $X(H)=1$. Suppose $B=\{H\}$. Determine $F_X(x \mid B)$.

Solution: From Example 3.2, $F_X(x)$ has the following form. We need $F_X(x \mid B)$ for all x.

For $x<0$, $\{X(\xi) \leq x\}=\phi$, so that $\{(X(\xi) \leq x) \cap B\}=\phi$, and $F_X(x \mid B)=0$.

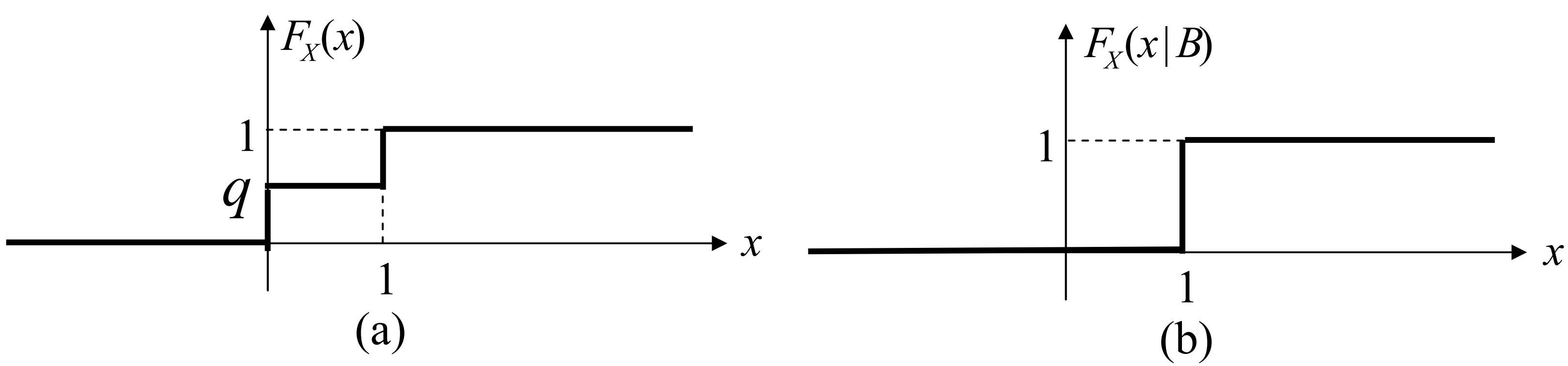


Fig. 4.3

For $0 \le x < 1$, $\{X(\xi) \le x\} = \{T\}$, so that

$\{(X(\xi) \le x) \cap B\} = \{T\} \cap \{H\} = \phi$ and $F_X(x \mid B) = 0.$

For $x \ge 1$, $\{X(\xi) \le x\} = \Omega$, and

$\{(X(\xi) \le x) \cap B\} = \Omega \cap \{B\} = \{B\}$ and $F_X(x \mid B) = \dfrac{P(B)}{P(B)} = 1$

(see Fig. 4.3(b)).

Example 4.5: Given $F_X(x)$, suppose $B = \{X(\xi) \le a\}$. Find $f_X(x \mid B)$.
Solution: We will first determine $F_X(x \mid B)$. From (4-11) and B as given above, we have

$$F_X(x \mid B) = \frac{P\{(X \le x) \cap (X \le a)\}}{P(X \le a)}. \tag{4-18}$$

For $x < a$, $(X \le x) \cap (X \le a) = (X \le x)$ so that

$$F_X(x \mid B) = \frac{P(X \le x)}{P(X \le a)} = \frac{F_X(x)}{F_X(a)}. \tag{4-19}$$

For $x \ge a$, $(X \le x) \cap (X \le a) = (X \le a)$ so that $F_X(x \mid B) = 1$.
Thus

$$F_X(x \mid B) = \begin{cases} \dfrac{F_X(x)}{F_X(a)}, & x < a, \\ 1, & x \ge a, \end{cases} \tag{4-20}$$

and hence

$$f_X(x \mid B) = \frac{d}{dx} F_X(x \mid B) = \begin{cases} \dfrac{f_X(x)}{F_X(a)}, & x < a, \\ 0, & \text{otherwise.} \end{cases} \tag{4-21}$$

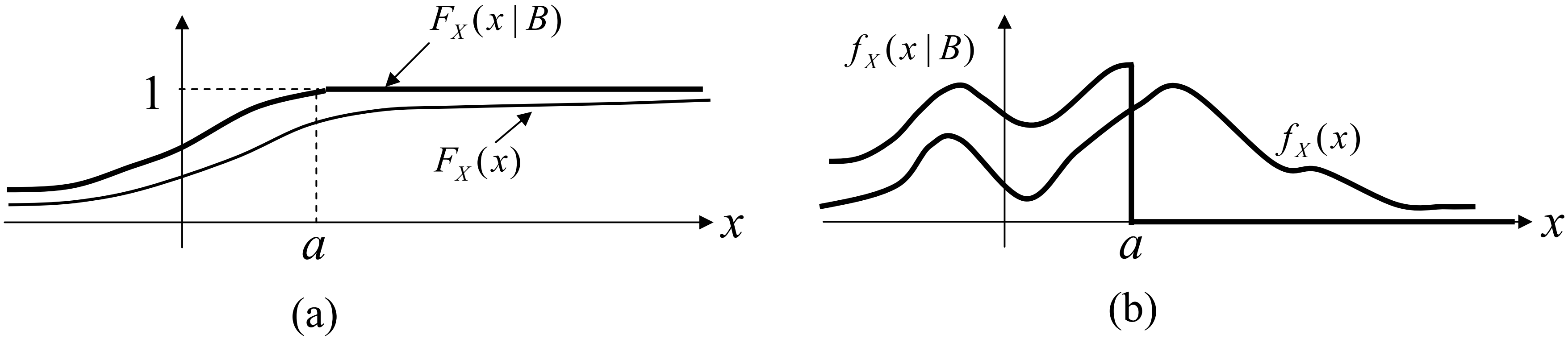


Fig. 4.4

Example 4.6: Let B represent the event $\{a < X(\xi) \le b\}$ with $b > a$. For a given $F_X(x)$, determine $F_X(x \mid B)$ and $f_X(x \mid B)$.

Solution:

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap (a < X(\xi) \le b)\}}{P(a < X(\xi) \le b)}$$

$$= \frac{P\{(X(\xi) \le x) \cap (a < X(\xi) \le b)\}}{F_X(b) - F_X(a)}. \qquad (4\text{-}22)$$

For $x < a$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \phi$, and hence $F_X(x \mid B) = 0.$ (4-23)

For $a \le x < b$, we have $\{ X(\xi) \le x \} \cap \{ a < X(\xi) \le b \} = \{ a < X(\xi) \le x \}$ and hence

$$F_X(x \mid B) = \frac{P\big(a < X(\xi) \le x\big)}{F_X(b) - F_X(a)} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}. \tag{4-24}$$

For $x \ge b$, we have $\{ X(\xi) \le x \} \cap \{ a < X(\xi) \le b \} = \{ a < X(\xi) \le b \}$ so that $F_X(x \mid B) = 1.$ (4-25)

Using (4-23)-(4-25), we get (see Fig. 4.5)

$$f_X(x \mid B) = \begin{cases} \dfrac{f_X(x)}{F_X(b) - F_X(a)}, & a < x \le b, \\ 0, & \text{otherwise.} \end{cases} \tag{4-26}$$

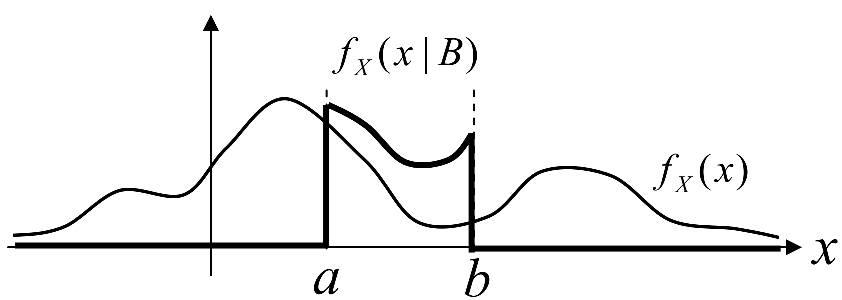


Fig. 4.5

We can use the conditional p.d.f together with the Bayes' theorem to update our a-priori knowledge about the probability of events in presence of new observations. Ideally, any new information should be used to update our knowledge. As we see in the next example, conditional p.d.f together with Bayes' theorem allow systematic updating. For any two events A and B, Bayes' theorem gives

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}. \tag{4-27}$$

Let $B = \{x_1 < X(\xi) \le x_2\}$ so that (4-27) becomes (see (4-13) and (4-17))

$$P\{A \mid (x_1 < X(\xi) \le x_2)\} = \frac{P\big((x_1 < X(\xi) \le x_2) \mid A\big)P(A)}{P\big(x_1 < X(\xi) \le x_2\big)}$$

$$= \frac{F_X(x_2 \mid A) - F_X(x_1 \mid A)}{F_X(x_2) - F_X(x_1)}P(A) = \frac{\int_{x_1}^{x_2} f_X(x \mid A)\,dx}{\int_{x_1}^{x_2} f_X(x)\,dx}P(A). \tag{4-28}$$

Further, let $x_1 = x,\ x_2 = x + \varepsilon,\ \varepsilon > 0,$ so that in the limit as $\varepsilon \to 0,$

$$\lim_{\varepsilon \to 0} P\{A \mid (x < X(\xi) \le x + \varepsilon)\} = P\big(A \mid X(\xi) = x\big) = \frac{f_X(x \mid A)}{f_X(x)} P(A). \tag{4-29}$$

or

$$f_{X|A}(x \mid A) = \frac{P(A \mid X = x) f_X(x)}{P(A)}. \tag{4-30}$$

From (4-30), we also get

$$P(A) \underbrace{\int_{-\infty}^{+\infty} f_X(x \mid A) dx}_{1} = \int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x) dx, \tag{4-31}$$

or

$$P(A) = \int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x) dx \tag{4-32}$$

and using this in (4-30), we get the desired result

$$f_{X|A}(x \mid A) = \frac{P(A \mid X = x) f_X(x)}{\int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x) dx}. \tag{4-33}$$

To illustrate the usefulness of this formulation, let us reexamine the coin tossing problem.

Example 4.7: Let $p = P(H)$ represent the probability of obtaining a head in a toss. For a given coin, a-priori p can possess any value in the interval (0,1). In the absence of any additional information, we may assume the a-priori p.d.f $f_P(p)$ to be a uniform distribution in that interval. Now suppose we actually perform an experiment of tossing the coin n times, and k heads are observed. This is new information. How can we update $f_P(p)$?

Solution: Let A= "k heads in n specific tosses". Since these tosses result in a specific sequence,

$$P(A \mid P = p) = p^k q^{n-k}, \qquad (4\text{-}34)$$

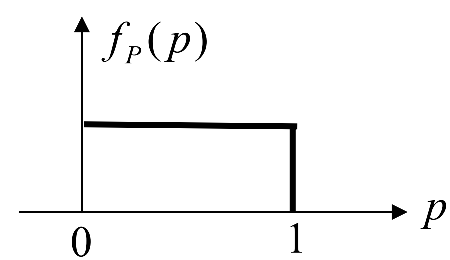


Fig.4.6

and using (4-32) we get

$$P(A) = \int_0^1 P(A \mid P = p) f_P(p)dp = \int_0^1 p^k (1-p)^{n-k} dp = \frac{(n-k)!\,k!}{(n+1)!}. \qquad (4\text{-}35)$$

The a-posteriori p.d.f $f_{P|A}(p \mid A)$ represents the updated information given the event A, and from (4-30)

$$f_{P|A}(p \mid A) = \frac{P(A \mid P = p) f_P(p)}{P(A)}$$

$$= \frac{(n+1)!}{(n-k)!\,k!} p^k q^{n-k}, \quad 0 < p < 1 \sim \beta(n,k). \qquad (4\text{-}36)$$

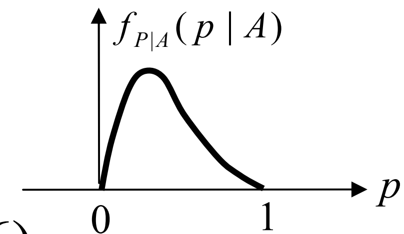


Fig. 4.7

Notice that the a-posteriori p.d.f of p in (4-36) is not a uniform distribution, but a beta distribution. We can use this a-posteriori p.d.f to make further predictions, For example, in the light of the above experiment, what can we say about the probability of a head occurring in the next $(n+1)$th toss?

Let B= "head occurring in the (n+1)th toss, given that k heads have occurred in n previous tosses".

Clearly $P(B \mid P = p) = p,$ and from (4-32)

$$P(B) = \int_0^1 P(B \mid P = p) f_P(p \mid A) dp. \tag{4-37}$$

Notice that unlike (4-32), we have used the a-posteriori p.d.f in (4-37) to reflect our knowledge about the experiment already performed. Using (4-36) in (4-37), we get

$$P(B) = \int_0^1 p \cdot \frac{(n+1)!}{(n-k)!\,k!} p^k q^{n-k} dp = \frac{k+1}{n+2}. \tag{4-38}$$

Thus, if n =10, and k = 6, then

$$P(B) = \frac{7}{12} = 0.58,$$

which is more realistic compare to p = 0.5.

To summarize, if the probability of an event X is unknown, one should make noncommittal judgement about its a-priori probability density function $f_X(x)$. Usually the uniform distribution is a reasonable assumption in the absence of any other information. Then experimental results (A) are obtained, and out knowledge about X must be updated reflecting this new information. Bayes' rule helps to obtain the a-posteriori p.d.f of X given A. From that point on, this a-posteriori p.d.f $f_{X|A}(x \mid A)$ should be used to make further predictions and calculations.

Stirling's Formula : What is it?

Stirling's formula gives an accurate approximation for $n!$ as follows:

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n} \qquad (4\text{-}39)$$

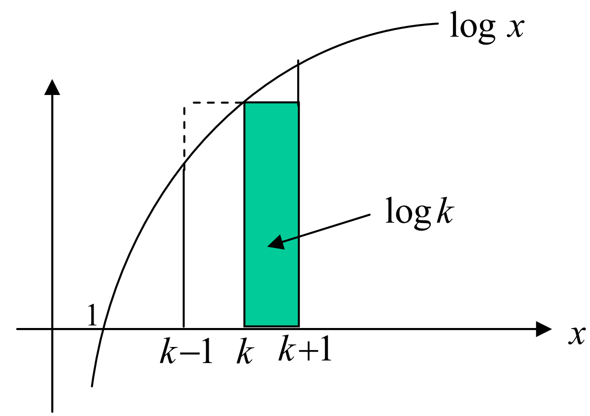
in the sense that the ratio of the two sides in (4-39) is near to one; i.e., their relative error is small, or the percentage error decreases steadily as n increases. The approximation is remarkably accurate even for small n. Thus $1! = 1$ is approximated as $\sqrt{2\pi}\,/\,e \simeq 0.9221,$ and $3! = 6$ is approximated as 5.836.

Prior to Stirling's work, DeMoivre had established the same formula in (4-39) in connection with binomial distributions in probability theory. However DeMoivre did not establish the constant term $\sqrt{2\pi}$ in (4-39); that was done by James Stirling ($\simeq$1730).

How to prove it?

We start with a simple observation: The function $\log x$ is a monotone increasing function, and hence we have

$$\int_{k-1}^{k} \log x\, dx < \log k < \int_{k}^{k+1} \log x\, dx.$$



Summing over $k = 1, 2, \cdots, n$ we get

$$\int_{0}^{n} \log x\, dx < \log n! < \int_{1}^{n+1} \log x\, dx$$

or $\left(\int \log x dx = x \log x - x\right)$

$$n \log n - n < \log n! < (n+1)\log(n+1) - n. \qquad (4\text{-}40)$$

The double inequality in (4-40) clearly suggests that $\log n!$ is close to the arithmetic mean of the two extreme numbers there. However the actual arithmetic mean is complicated and it involves several terms. Since $(n+\frac{1}{2})\log n - n$ is quite close to the above arithmetic mean, we consider the difference[1]

$$a_n \triangleq \log n! - (n+\tfrac{1}{2})\log n + n. \tag{4-41}$$

This gives

$$\begin{aligned} a_n - a_{n+1} &= \log n! - (n+\tfrac{1}{2})\log n + n - \log(n+1)! \\ &\quad + (n+\tfrac{3}{2})\log(n+1) - (n+1) \\ &= -\log(n+1) - (n+\tfrac{1}{2})\log n + (n+\tfrac{3}{2})\log(n+1) - 1 \\ &= (n+\tfrac{1}{2})\log\tfrac{n+1}{n} - 1 = (n+\tfrac{1}{2})\log\frac{n+\frac{1}{2}+\frac{1}{2}}{n+\frac{1}{2}-\frac{1}{2}} - 1. \end{aligned}$$

[1]According to W. Feller this clever idea to use the approximate mean $(n+\frac{1}{2})\log n - n$ is due to H.E. Robbins, and it leads to an elementary proof.

Hence[1]

$$a_n - a_{n+1} = (n+\tfrac{1}{2})\log\left(\frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}}\right) - 1$$

$$= 2(n+\tfrac{1}{2})\left(\frac{1}{2n+1} + \frac{1}{3(2n+1)^3} + \frac{1}{5(2n+1)^5} + \cdots\right) - 1$$

$$= \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \frac{1}{7(2n+1)^6} + \cdots > 0 \qquad (4\text{-}42)$$

Thus $\{a_n\}$ is a monotone decreasing sequence and let c represent its limit, i.e.,

$$\lim_{n\to\infty} a_n = c \qquad (4\text{-}43)$$

From (4-41), as $n \to \infty$ this is equivalent to

$$n! \sim e^c n^{n+\frac{1}{2}} e^{-n}. \qquad (4\text{-}44)$$

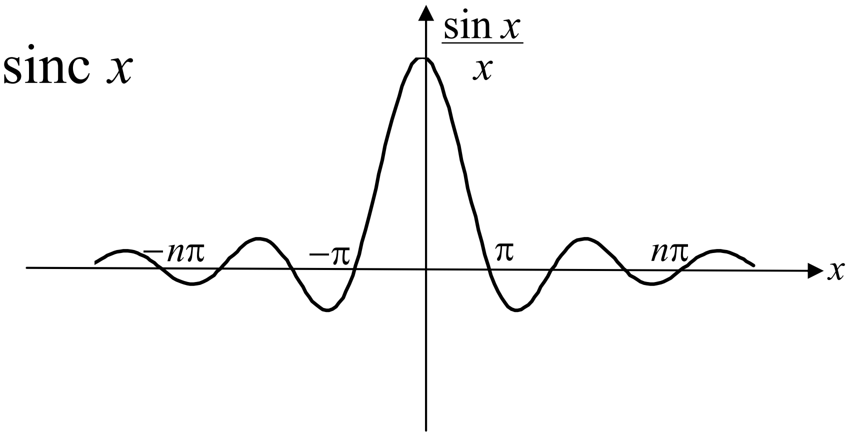
To find the constant term c in (4-44), we can make use of a formula due to Wallis ($\simeq$ 1655).

---

[1]By Taylor series expansion

$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots;\ \log\frac{1}{(1-x)} = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \frac{1}{4}x^4 + \cdots$

The well known function $\frac{\sin x}{x} = \text{sinc } x$ goes to zero at $x = \pm n\pi$; moreover these are the *only* zeros of this function. Also $\frac{\sin x}{x}$ has no finite poles. (All poles are at infinity). As a result we can write



$$\frac{\sin x}{x} = (1-\tfrac{x^2}{\pi^2})(1-\tfrac{x^2}{4\pi^2})\cdots(1-\tfrac{x^2}{n^2\pi^2})\cdots$$

or [for a proof of this formula, see chapter 4 of Dienes, *The Taylor Series*]

$$\sin x = x(1-\tfrac{x^2}{\pi^2})(1-\tfrac{x^2}{4\pi^2})\cdots(1-\tfrac{x^2}{n^2\pi^2})\cdots,$$

which for $x = \pi/2$ gives the Wallis' formula

$$1 = \frac{\pi}{2}(1-\tfrac{1}{2^2})(1-\tfrac{1}{4^2})\cdots(1-\tfrac{1}{(2n)^2}) = \frac{\pi}{2}(\tfrac{1\cdot 3}{2^2})(\tfrac{3\cdot 5}{4^2})(\tfrac{5\cdot 7}{6^2})\cdots(\tfrac{(2n-1)\cdot(2n+1)}{(2n)^2})\cdots$$

or
$$\frac{\pi}{2} = \prod_{n=1}^{\infty} \frac{4n^2}{(2n-1)(2n+1)} = (\tfrac{2^2}{1\cdot 3})(\tfrac{4^2}{3\cdot 5})(\tfrac{6^2}{5\cdot 7})\cdots(\tfrac{(2n)^2}{(2n-1)\cdot(2n+1)})\cdots$$
$$= (\tfrac{2\cdot 4\cdot 6\cdots 2n\cdots}{1\cdot 3\cdot 5\cdots(2n-1)\cdots})^2 \tfrac{1}{2n+1}\cdots.$$

Thus as $n \to \infty$, this gives
$$\sqrt{\pi} = \frac{2\cdot 4\cdot 6\cdots 2n}{1\cdot 3\cdot 5\cdots(2n-1)} \frac{1}{\sqrt{n+\frac{1}{2}}}\cdots$$
$$= \frac{(2\cdot 4\cdot 6\cdots 2n)^2}{(2n)!} \frac{1}{\sqrt{n+\frac{1}{2}}}\cdots = 2^{2n}\frac{(n!)^2}{(2n)!}\frac{1}{\sqrt{n+\frac{1}{2}}}.$$

Thus as $n \to \infty$
$$\log\sqrt{\pi} = 2n\log 2 + 2\log n! - \log(2n)! - \frac{1}{2}\log(n+\tfrac{1}{2}) \quad (4\text{-}45)$$

But from (4-41) and (4-43)
$$\lim_{n\to\infty}\log n! = c + \lim_{n\to\infty}\{(n+\tfrac{1}{2})\log n - n\} \quad (4\text{-}46)$$

and hence letting $n \to \infty$ in (4-45) and making use of (4-46) we get

$$\log\sqrt{\pi} = c - \tfrac{1}{2}\log 2 - \lim_{n\to\infty} \tfrac{1}{2}\log(1+\tfrac{1}{2n}) = c - \tfrac{1}{2}\log 2$$

which gives

$$e^c = \sqrt{2\pi}. \tag{4-47}$$

With (4-47) in (4-44) we obtain (4-39), and this proves the Stirling's formula.

Upper and Lower Bounds

It is possible to obtain reasonably good upper and lower bounds for $n!$ by elementary reasoning as well. To see this, note that from (4-42) we get

$$a_n - a_{n+1} < \frac{1}{3}\left(\frac{1}{(2n+1)^2} + \frac{1}{(2n+1)^4}\cdots\right)$$

$$= \frac{1}{3[(2n+1)^2-1]} = \frac{1}{12n(n+1)} = \frac{1}{12n} - \frac{1}{12(n+1)}$$

so that $\{a_n - 1/12n\}$ is a monotonically increasing sequence whose limit is also c. Hence for any finite n

$$a_n - \frac{1}{12n} < c \quad \Rightarrow \quad a_n < c + \frac{1}{12n}$$

and together with (4-41) and (4-47) this gives

$$n! < \sqrt{2\pi}\ n^{n+\frac{1}{2}} e^{-n\left(1-\frac{1}{12n^2}\right)} \tag{4-48}$$

Similarly from (4-42) we also have

$$a_n - a_{n+1} > \frac{1}{3(2n+1)^2} > \frac{1}{12n+1} - \frac{1}{12(n+1)+1} > 0$$

so that $\{a_n - 1/(12n+1)\}$ is a monotone decreasing sequence whose limit also equals c. Hence

$$a_n - \frac{1}{12n+1} > c \quad \Rightarrow \quad a_n > c + \frac{1}{12n+1}$$

or

$$n! > \sqrt{2\pi}\ n^{n+\frac{1}{2}} e^{-n\left(1-\frac{1}{12n(n+1)}\right)}. \tag{4-49}$$

Together with (4-48)-(4-49) we obtain

$$\sqrt{2\pi}\; n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi}\; n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n}}. \qquad (4\text{-}50)$$

Stirling's formula also follows from the asymptotic expansion for the Gamma function given by

$$\Gamma(x) = \sqrt{2\pi}\; e^{-x} x^{x-\frac{1}{2}} \left\{ 1 + \frac{1}{12x} + \frac{1}{286x^2} - \frac{139}{51840x^3} + o\left(\frac{1}{x^4}\right) \right\} \qquad (4\text{-}51)$$

Together with $\Gamma(x+1) = x\Gamma(x)$, the above expansion can be used to compute numerical values for real x.
For a derivation of (4-51), one may look into Chapter 2 of the classic text by Whittaker and Watson (*Modern Analysis*).

We can use Stirling's formula to obtain yet another approximation to the binomial probability mass function. Since

$$\binom{n}{k} p^k q^{n-k} = \frac{n!}{(n-k)!\,k!} p^k q^{n-k}, \qquad \text{(4-52)}$$

using (4-50) on the right side of (4-52) we obtain

$$\binom{n}{k} p^k q^{n-k} > c_1 \sqrt{\frac{n}{2\pi (n-k)k}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$$

and

$$\binom{n}{k} p^k q^{n-k} < c_2 \sqrt{\frac{n}{2\pi (n-k)k}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$$

where

$$c_1 = e^{\left\{\frac{1}{12n+1} - \frac{1}{12(n-k)} - \frac{1}{12k}\right\}}$$

and

$$c_2 = e^{\left\{\frac{1}{12n} - \frac{1}{12(n-k)+1} - \frac{1}{12k+1}\right\}}.$$

Notice that the constants c_1 and c_2 are quite close to each other.