13 Bernoulli Experiment and Its Related Distributions

An experiment is called a **Bernoulli experiment** if there are only two possible outcome: success with probability p and failure with probability $(1-p)$ where $0 < p < 1$.

We say X is a Bernoulli random variable or $p(x)$ is **Bernoulli distribution** $B(p)$ if

$$p(x) = \begin{cases} p \quad \text{if} \quad x = 1, \\ 1-p \quad \text{if} \quad x = 0. \end{cases}$$

Here $X = 1$ represents the outcome is "success" and $X = 0$ represents the outcome is "failure".

13.1 Bernoulli Distribution

Proposition 1. *The mean μ and variance σ^2 of the Bernoulli distribution are given by p and $(p - p^2)$ respectively.*

Proof. Since

$$\mu = E(X) = 0 \cdot p(0) + 1 \cdot p(1) = p$$

we have

$$E(X^2) = 0^2 \cdot p(0) + 1^2 \cdot p(1) = p$$

and

$$\sigma^2 = E(X^2) - \mu^2 = p - p^2.$$

□

Example 1. The number of head X obtained in tossing a fair coin is a Bernoulli experiment with $p = 0.5$ and probability distribution:

$$p(x) = \begin{cases} 0.5 & \text{if} \quad x = 1, \\ 0.5 & \text{if} \quad x = 0. \end{cases}$$

Moreover, **the expected number of head** is 0.5 and the variance is

$$0.5 - 0.5^2 = 0.25.$$

13.2 Geometric Distribution

One may also think of the following interesting situation. We perform a series of Bernoulli experiments (tossing a coin) until we get the first success (first head) then we stop. Each Bernoulli trial is **independent**.

Independent trial means that one occurrence (or non-occurrence) of an event does not influence the successive occurrences or non-occurrence of that event.

What is the probability $p(x)$ that we get the first success (the first head) in the xth experiment?

Suppose the probability of success is p, then in the first $(x-1)$ trials, we must get $(x-1)$ failures (tails), the probability is $(1-p)^{x-1}$.

At the xth trial we must get a success (head) hence the probability is for getting the first head at the xth trial is

$$\boxed{p(x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots.}$$

We note

$$p(x) \geq 0$$

and

$$\sum_{x=1}^{\infty} p(x) = 1.$$

Thus $p(x)$ is a discrete distribution and it is called the **geometric distribution** $Geo(p)$.

Example 2. In an one-machine production system, everyday the machine has a probability of 0.01 to be broken. Find the probability that the machine can survive over one month (more than 30 days).

The probability that the machine breaks down on the xth day is

$$p(x) = 0.01 \cdot (0.99)^{x-1}, \quad x = 1, 2, 3, \ldots .$$

The probability that it will break down within one month is

$$0.01 \sum_{x=1}^{30} (0.99)^{x-1} = 1 - 0.99^{30}.$$

Thus the probability that the machine can survive over one month will be

$$1 - (1 - 0.99^{30}) = 0.99^{30} = 0.7397.$$

Proposition 2. *The mean* μ *and variance* σ^2 *of the geometric distribution are given by* $1/p$ *and* $(1-p)/p^2$ *respectively.*

Proof. Before we show the results, we need the following formula.

For $|y| < 1$, we have

$$\begin{cases} \dfrac{1}{1-y} = \displaystyle\sum_{k=0}^{\infty} y^k. \\ \dfrac{1}{(1-y)^2} = \displaystyle\sum_{k=1}^{\infty} k y^{k-1}. \\ \dfrac{2y}{(1-y)^3} = \displaystyle\sum_{k=2}^{\infty} k(k-1) y^{k-1}. \end{cases}$$

NOTE: For $|y| < 1$, we have $1 + y + y^2 + \cdots + = \frac{1}{1-y}$.

Let

$$B = 1 + 2y + 3y^2 + \ldots + \quad \text{then} \quad yB = y + 2y^2 + 3y^3 + \ldots + .$$

We have

$$(1-y)B = 1 + y + y^2 + \ldots + = \frac{1}{1-y}.$$

Hence

$$B = \frac{1}{(1-y)^2}.$$

Let

$$A = (2)(1)y + (3)(2)y^2 + (4)(3)y^3 + (5)(4)y^4 + \ldots +$$

Then

$$yA = (2)(1)y^2 + (3)(2)y^3 + (4)(3)y^4 + (5)(4)y^5 + \ldots +$$

and

$$(1-y)A = (2)(1)y + (2)(2)y^2 + (2)(3)y^3 + (2)(4)y^4 + \ldots + .$$

Therefore

$$A = \frac{2y}{1-y}\left\{1 + 2y + 3y^2 + 4y^3 + \ldots +\right\} = \frac{2y}{1-y} \times \frac{1}{(1-y)^2} = \frac{2y}{(1-y)^3}.$$

We have

$$\begin{aligned}
\mu &= E(X) \\
&= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\
&= p\sum_{x=1}^{\infty} x(1-p)^{x-1} = \frac{p}{p^2} = \frac{1}{p}
\end{aligned}$$

and

$$\begin{aligned}
E(X^2) - E(X) &= E(X(X-1)) \\
&= \sum_{x=2}^{\infty} x(x-1)p(1-p)^{x-1} \\
&= p\sum_{x=2}^{\infty} x(x-1)(1-p)^{x-1} = p \cdot \frac{2(1-p)}{p^3} = \frac{2(1-p)}{p^2}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\sigma^2 &= E(X^2) - \mu^2 = E(X(X-1)) + \mu - \mu^2 \\
&= (\frac{2}{p^2} - \frac{2}{p}) + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.
\end{aligned}$$

□

13.3 Binomial Distribution

One more interesting situation is to obtain x successful trials in n Bernoulli trials. In fact, an experiment that satisfies the following condition is called a **binomial experiment**. There are n identical independent Bernoulli trials. In other words, the given experiment is repeated n times. All these repetitions are performed under **identical conditions**.

The random variable x that represents the number of successes in n trials for a binomial experiment is called a **binomial random variable**. The probability distribution of X in such experiments is called the binomial probability distribution or simply **binomial distribution** $Bin(n, p)$.

In a binomial experiment, the probability of exactly x **successes** in n trials is given by the binomial formula

$$\boxed{p(x) = \binom{n}{x} p^x (1-p)^{n-x}}$$

where n, total number of trials; p, probability of a success; $1-p$, probability of a failure; x, number of successes in n trials; $n-x$, number of failures in n trials.

We note that

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \geq 0$$

and

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-p))^n = 1.^1$$

Thus we see that $p(x)$ is really a probability distribution of discrete type.

Proposition 3. *The mean and variance of the binomial distribution are*

$$\mu = np \quad \textit{and} \quad \sigma^2 = npq.$$

Here $q = 1 - p$.

Exercise 1. To show the above proposition. You may write $X = X_1 + X_2 + \cdots + X_n$ where X_i are independent Bernoulli random variables. [2] We recall that

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

and

$$Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n).$$

---

[1] $(a+b)^n = \sum_{r=0}^{n} nCr \cdot a^r b^{n-r}$.

[2] If X follows $Bin(n,p)$ and Y follows $Bin(m,p)$ and they are independent then $X+Y$ follows $Bin(m+n,p)$.

Example 3. A fair coin is toss $2n$ times.
(a)What is the probability that number of heads equal to number of tails?
(b) What happen when n is very large?

(a) We employ the binomial distribution with $p = q = 0.5$. Then the probability will be given by

$$\boxed{p_n = \frac{(2n)!}{n!n!}\left(\frac{1}{2}\right)^{2n}.}$$

(b) For large n, to analyze the situation, we can employ the **Stirling formula** in Page 5 of Part 2:

$$n! \approx \sqrt{2n\pi}\left(\frac{n}{e}\right)^n.$$

Then we get

$$\boxed{p_n \approx \frac{2^{2n}}{\sqrt{\pi n}}(\frac{1}{2})^{2n} = \frac{1}{\sqrt{\pi n}}.}$$

Thus p_n is getting smaller and smaller to zero when n increases.

13.4 Some Computational Issue

For the **Binomial distribution** $Bin(n, p)$, we have

$$p_r = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

and

$$p_{r-1} = \frac{n!}{(r-1)!(n-r+1)!} p^{r-1} (1-p)^{n-r+1}.$$

Then we have

$$\boxed{p_r = \frac{(n-r+1)p}{r(1-p)} p_{r-1} \quad r = 1, 2, \ldots, n.} \tag{13.1}$$

Begin with

$$p_0 = (1-p)^n$$

we can compute $p_1, p_2, \ldots, p_n$ by using Equation (13.1). This recursive form is useful when we are asked to compute the probability with n and r being very large.

13.5 Negative Binomial Distribution

Sometimes we are not just interested in obtaining the first success in the kth Bernoulli trial like the geometric distribution. But we are interested in obtaining the rth success in the kth trial ($k \geq r \geq 1$).

Now suppose that the probability of success in each independent Bernoulli trial is p. What is the probability

$$B_r(k, p) \quad \text{where} \quad k \geq r$$

that the rth success is obtained in the kth trial?

To obtain $B_r(k, p)$, we note that we must have **a success in the kth trial** and at the same time we must have $(r-1)$ **successes in the first** $(k-1)$ **trials**.

The former probability is p and the latter probability is

$$\binom{k-1}{r-1} p^{r-1}(1-p)^{k-r}.$$

Since they are independent events so the required probability is

$$B_r(k,p) = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} \times p, \quad 1 \le r \le k, \quad k = r, r+1, \ldots.$$

We note that

$$B_r(k,p) = \binom{k-1}{r-1} p^{r}(1-p)^{k-r} \ge 0$$

and

$$\sum_{k=r}^{\infty} B_r(k,p) = 1.$$

Thus $B_r(k,p)$ is a distribution of discrete type and is called the **Negative Binomial Distribution** $B_r(p)$. We see that when r is 1, $B_1(k,p)$ is the geometric distribution in k. Thus the negative binomial distribution is a generalization of the geometric distribution.

Proposition 4. *The mean μ and variance σ^2 of the Negative Binomial Distribution $B_r(k, p)$ are r/p and $r(1-p)/p^2$ respectively.*

Proof. Recall that the mean and variance of a geometric distributed variable are $1/p$ and $(1-p)/p^2$ respectively. To obtain the rth success it is equivalent to perform the geometric process r times. Suppose Y is the number of trials to obtain the rth success in a sequence of Bernoulli trials, then

$$Y = X_1 + X_2 + \ldots + X_r,$$

where x_i is the number of the trials to obtain the ith success immediately after the $(i-1)$th success for $i = 1, 2, \ldots, r$. Thus all the x_i are independent Geometric distributed random variables. Hence we have

$$\mu = E(Y) = E(X_1) + E(X_2) + \ldots + E(X_r) = \frac{1}{p} + \ldots + \frac{1}{p} = \frac{r}{p}$$

and

$$\begin{aligned} \sigma^2 &= \mathrm{Var}(Y) = Var(X_1) + Var(X_2) + \ldots + Var(X_r) \\ &= \frac{1-p}{p^2} + \ldots + \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}. \end{aligned}$$

□

14 Poisson Distribution

A **Poisson random variable** X with parameter λ has the probability function

$$\boxed{P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}; \quad k=0,1,2,\ldots.}$$

Exercise 2. For the above random variable which follows the Poisson distribution $Poi(\lambda)$, show that we have

$$E(X)=Var(X)=\lambda.$$

Hint: we note that

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{\lambda} \quad \text{and} \quad \sum_{k=0}^{\infty}k\times\frac{\lambda^k}{k!}=\sum_{k=1}^{\infty}\lambda\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{\lambda}$$

and

$$\sum_{k=0}^{\infty}k(k-1)\times\frac{\lambda^k}{k!}=\sum_{k=2}^{\infty}\lambda^2\frac{\lambda^{k-2}}{(k-2)!}=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=\lambda^2e^{\lambda}.$$

Then we have $Var(X)=E(X^2)-E(X)^2=E(X(X-1))+E(X)-E(X)^2$. [3]

[3] $E(X(X-1))=\sum_{x=0}^{\infty}x(x-1)p(x)=\sum_{x=0}^{\infty}(x^2-x)p(x)=\sum_{x=0}^{\infty}x^2p(x)-\sum_{x=0}^{\infty}xp(x)=E(X^2)-E(X)$.

Example 4. The arrival process of customers follows a Poisson distribution with mean 2 per hour. Find the probability of having at least one arrival of customer in one hour.

The probability of having no customer in one hour is $p_0 = e^{-2} \cdot \frac{2^0}{0!}$. Therefore the probability of having at least one customer will be

$$1 - p_0 = 1 - e^{-2} = 0.8647.$$

Remark: For the **Poisson distribution** $Poi(\lambda)$, we have

$$p_r = \frac{\lambda^r}{r!} e^{-\lambda} \quad \text{and} \quad p_{r-1} = \frac{\lambda^{r-1}}{(r-1)!} e^{-\lambda}.$$

Then we have

$$\boxed{p_r = \frac{\lambda}{r} p_{r-1}, \quad r = 1, 2, \ldots, n.} \tag{14.1}$$

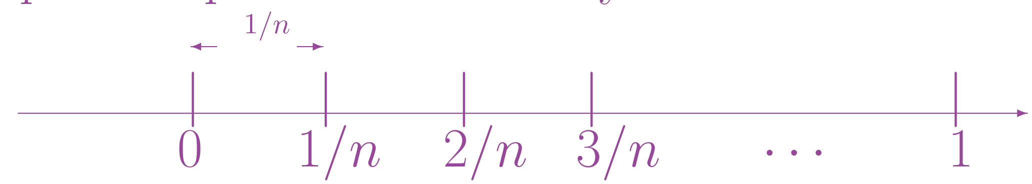
Begin with

$$p_0 = e^{-\lambda}$$

we can compute $p_1, p_2, \ldots, p_n$ by using Equation (14.1).

14.1 From Binomial to Poisson*

This part is optional and is for your information only.



The Poisson Process.

• We divide the unit time interval into n equal sub-intervals of length $1/n$. We assume that in each sub-interval, with probability p that there will be an occurrence (e.g. an arrival of customer or an accident). And with probability $(1-p)$ there is no occurrence. Therefore it is a Bernoulli process in each sub-interval.

• We further assume that the probability p is proportional to the length of the sub-interval, (i.e. $p \propto 1/n$) with a positive constant λ. Thus we have $p = \lambda \times (1/n)$. [4]

• Assuming independence among all the sub-intervals, the probability of having exactly k occurrence will follow $Bin(n,p)$:

$$P\{X=k\} = \frac{n!}{(n-k)!k!}p^k(1-p)^{n-k} = \frac{n!}{(n-k)!k!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}.$$

[4] $\lambda = np$ is the mean of the binomial distribution.

Suppose we are observing the arrival of customers 1 **hour** outside a bank.

• Assume that $n = 60$, i.e., we observe the arrival process in every minute (length of each interval is 1 minute). We further assume that $\lambda = 5$. Then the probability of having an arrival in each of the sub-interval is $\lambda/n = 5/60 = 1/12$ and no customer will be $11/12$. Then the probability of having $k = 2$ arrived customers in 1 hour is

$$P\{X = 2\} = \frac{60!}{58!2!}(\frac{1}{12})^2(\frac{11}{12})^{58} = 0.079045.$$

• Suppose 1 minute is too long, we take $n = 3600$, i.e., we observe the arrival process in every second (length of each interval is 1 second). Then the probability of having an arrival in each of the sub-interval is $\lambda/n = 5/3600 = 1/720$ and no customer will be $719/720$. Then the probability of having $k = 2$ arrived customers in one hour is

$$P\{X = 2\} = \frac{3600!}{3598!2!}(\frac{1}{720})^2(\frac{719}{720})^{3598} = 0.084142.$$

• Finally if $\lambda = 5$ and $n \to \infty$ then the probability of having 2 arrived customer in one hour can be computed by using the **Poisson distribution** $Poi(5)$

$$P\{X = 2\} = e^{-5}\frac{5^2}{2!} = 0.084224.$$

One may "derive" the Poisson distribution from the Binomial distribution by letting $\lambda = np$ and $n \to \infty$. We derive the relationship as follows:

$$
\begin{aligned}
P\{X = k\} &= \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= \frac{1}{k!} \left(p \cdot (n-k+1)\right) \left(p \cdot (n-k+2)\right) \ldots \left(p \cdot (n)\right) (1-p)^{n-k} \\
&= \frac{1}{k!} \left(\frac{(n-k+1)\lambda}{n}\right) \left(\frac{(n-k+2)\lambda}{n}\right) \ldots (\lambda) \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{as } n \to \infty.
\end{aligned}
$$

Remark 1. We have for a fixed k,

$$\frac{(n-k+1)\lambda}{n} \to \lambda \quad \text{as} \quad n \to \infty.$$

Remark 2. We have

$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda} \quad \text{as} \quad n \to \infty.$$

Remark 3. *Suppose X_1 and X_2 are* **two independent Poisson random variables** *follow $Poi(\lambda_1)$ and $Poi(\lambda_2)$ respectively. Then $X = X_1 + X_2$ is again* **a Poisson random variable** *follows $Poi(\lambda_1 + \lambda_2)$.*

$$
\begin{aligned}
P(X = y) &= \sum_{k=0}^{y} P(X_1 = k) \cdot P(X_2 = y - k) \\
&= \sum_{k=0}^{y} e^{-\lambda_1} \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_2} \frac{\lambda_2^{y-k}}{(y-k)!} \\
&= e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{y} \frac{\lambda_1^k}{k!} \cdot \frac{\lambda_2^{y-k}}{(y-k)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{y!} \sum_{k=0}^{y} \frac{y!}{k!(y-k)!} \cdot \lambda_1^k \cdot \lambda_2^{y-k} = \frac{e^{-(\lambda_1+\lambda_2)}}{y!} (\lambda_1 + \lambda_2)^y.
\end{aligned}
$$

Thus X follows $Poi(\lambda_1 + \lambda_2)$. In general, if X_i follows $Poi(\lambda_i)$ $(i = 1, 2, \ldots, n)$ and they are independent then we have

$$
X_1 + X_2 + \cdots + X_n \quad \text{follows} \quad Poi(\lambda_1 + \lambda_2 + \cdots + \lambda_n).
$$

15 Normal Distribution

The **Normal distribution** has a continuous probability density function taking the following form:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad -\infty < x < \infty.$$

Proposition 5. *The mean and variance of the above normal distribution are μ and σ^2 respectively.*

Proof. We note that if we apply integration by substitution $y = \frac{x-\mu}{\sigma}$,

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

$$= \underbrace{\sigma \int_{-\infty}^{\infty} \frac{y}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy}_{=0} + \underbrace{\mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy}_{=\mu} .^{5}$$

Thus we have

$$E(X) = \mu.$$

[5] In this course, we assume without proof that $\int_{-\infty}^{\infty} e^{\frac{-y^2}{2}} dy = \sqrt{2\pi}$.

We then consider the same substitution,

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

Thus we have

$$E(X^2) = \underbrace{\sigma^2 \int_{-\infty}^{\infty} \frac{y^2}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy}_{=\sigma^2} + \underbrace{2\sigma\mu \int_{-\infty}^{\infty} \frac{y}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy}_{=0} + \underbrace{\mu^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy}_{=\mu^2}.$$

For the first term we note that

$$\int \frac{y \cdot y}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy = \int \frac{y}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} d(\frac{y^2}{2}) = -\int \frac{y}{\sqrt{2\pi}} de^{\frac{-y^2}{2}} = \underbrace{\frac{1}{\sqrt{2\pi}} \left(\int e^{\frac{-y^2}{2}} dy - y e^{\frac{-y^2}{2}} \right)}_{\text{integration by part}}.$$

$$\int_{-\infty}^{\infty} \frac{y^2}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} (\underbrace{\int_{-\infty}^{\infty} e^{\frac{-y^2}{2}} dy}_{\sqrt{2\pi}} - \underbrace{y e^{\frac{-y^2}{2}} |_{-\infty}^{\infty}}_{=0}) = 1.$$

Finally we have

$$Var(X) = E(X^2) - E(X)^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

□

15.1 From Binomial to Normal*

This section is optional and it presents the relation between the Binomial distribution and the Normal distribution.

- Consider the Binomial distribution $Bin(n, p)$ for large n and $p \approx 0.5$, i.e. $q = 1 - p \approx 0.5$.

- Its mean and variance are given, respectively, by

$$\mu = np \quad \text{and} \quad \sigma^2 = np(1-p) = npq$$

where $q = 1 - p$.

We have

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}, \quad x = 0, 1, 2, \cdots, n.$$

Let $y = x - np$ be **small perturbation** from its mean $\mu = np$

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{n!}{(y+np)!(nq-y)!} p^{np+y} q^{nq-y}.$$

Now we recall the Stirling's formula for large n:

$$n! \approx \sqrt{2n\pi}\left(\frac{n}{e}\right)^n.$$

- Then we have

$$(y+np)! \approx \sqrt{2(y+np)\pi}\left(\frac{y+np}{e}\right)^{y+np}$$

and

$$(nq-y)! \approx \sqrt{2(nq-y)\pi}\left(\frac{nq-y}{e}\right)^{nq-y}.$$

We have

$$P(x) \approx \sqrt{\frac{n}{2\pi(np+y)(nq-y)}} \times \frac{n^n p^{np+y} q^{nq-y}}{(np+y)^{np+y}(nq-y)^{nq-y}}.$$

- We note that

$$\sqrt{\frac{n}{2\pi(np+y)(nq-y)}} \approx \sqrt{\frac{1}{2\pi npq}} \tag{15.1}$$

and

$$\frac{n^n p^{np+y} q^{nq-y}}{(np+y)^{np+y}(nq-y)^{nq-y}} = \left(\frac{np}{np+y}\right)^{np+y}\left(\frac{nq}{nq-y}\right)^{nq-y} = \frac{1}{K}.$$

- Now we note that

$$K = \left(1+\frac{y}{np}\right)^{np+y}\left(1-\frac{y}{nq}\right)^{nq-y}$$

and we have

$$\begin{aligned}
\log_e K &= (np+y)\log_e\left(1+\tfrac{y}{np}\right) + (nq-y)\log_e\left(1-\tfrac{y}{nq}\right)\\
&\approx (np+y)\left(\tfrac{y}{np} - \tfrac{y^2}{2n^2p^2} + \tfrac{y^3}{3n^3p^3}\right)\\
&\quad +(nq-y)\left(-\tfrac{y}{nq} - \tfrac{y^2}{2n^2q^2} - \tfrac{y^3}{3n^3q^3}\right)
\end{aligned}$$

because for small z, $|z| < 1$ we have

$$\boxed{\log_e(1+z) \approx z - \frac{z^2}{2} + \frac{z^3}{3}}$$

We have

$$\log_e K \approx \boxed{\frac{y^2}{2n}(\frac{1}{p}+\frac{1}{q})} + \frac{y^3}{6n^2}(\frac{1}{q^2}-\frac{1}{p^2}) + \text{higher power of } \frac{1}{n^2}.$$

Hence for large n, we keep the first term and drop the others. Then we obtain

$$K \approx e^{\frac{y^2}{2npq}} \quad \text{or} \quad \frac{1}{K} \approx e^{-\frac{y^2}{2npq}}. \tag{15.2}$$

Combining the results in Equations (15.1) and (15.2) we have (using the parameters of $Bin(n,p)$) :

$$P(x) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{y^2}{2npq}}.$$

Replace npq by σ^2, np by μ and y by $x - np = x - \mu$, we have (using the parameters of mean μ and variance σ^2):

$$P(x) \approx \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}.$$

This is the normal distribution $N(\mu, \sigma^2)$.

15.2 Properties of a Normal Distribution

1. Usually we denote a normal distribution of mean μ and variance σ^2 by $N(\mu, \sigma^2)$.
2. We note that the probability function
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$
is symmetric at the mean $x = \mu$.
3. The **mean, mode and the median** of the distribution are equal to μ.
4. The C.D.F. is
$$\boxed{\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx.}$$
When $\mu = 0$, we have for $t \geq 0$
$$\boxed{\Phi(-t) = 1 - \Phi(t).}$$
5. Suppose Z_1 and Z_2 are **two independent normal random variables** follow $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively. Then $Z = Z_1 + Z_2$ is **a normal random variable** follows $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. We shall assume this is true without a proof.

15.3 Standard Normal Distribution

Suppose the random variable X follows $N(\mu, \sigma^2)$, i.e., the p.d.f. of X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad -\infty < x < \infty.$$

We then consider a new random variable

$$\boxed{Z = \frac{X-\mu}{\sigma}.}$$

• For this z we note that its mean is

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{E(X)-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

and its variance is

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X-\mu) = \frac{\sigma^2}{\sigma^2} = 1.$$

• After this transform, we have another normal random variable whose mean is shifted to 0 and its variance is scaled to 1.

• What is the probability density function of Z?

We note that

$$P(Z \leq z) = P\left(\frac{X-\mu}{\sigma} \leq z\right) = P(X \leq \sigma z + \mu) = \int_{-\infty}^{\sigma z+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx.$$

Now by considering the substitution

$$y = \frac{x-\mu}{\sigma}$$

we have

$$\frac{dy}{dx} = \frac{1}{\sigma}$$

and

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

which is the C.D.F. of $N(0,1)$ which is called the **standard normal distribution**. Thus any normal distribution $N(\mu, \sigma^2)$ can be converted to the standard normal distribution $N(0,1)$. And the following table gives the probabilities:

$$P(Z \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

The following table gives the probability $P(Z \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

z	.0000	.0100	.0200	.0300	.0400	.0500	.0600	.0700	.0800	.0900
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

Example 5. Let X follows the normal distribution $N(2, 2^2)$, find the probabilities:
(a) $P(X \leq 3)$; (b) $P(1 \leq X)$; (c) $P(1 \leq X \leq 3)$.

We have that $\mu = 2$ and $\sigma = 2$.
(a) We note that $X \leq 3$ is equivalent to $Z = \frac{X-2}{2} \leq \frac{3-2}{2} = 0.5$.
Thus $X \leq 3$ is equivalent to $Z \leq 0.5$. Then from the table we have

$$P(X \leq 3) = P(Z \leq 0.5) = \Phi(0.5) = 0.6915.$$

(b) We note that $1 \leq X$ is equivalent to $-0.5 = \frac{1-2}{2} \leq \frac{x-2}{2} = Z$.
Thus $1 \leq X$ is equivalent to $-0.5 \leq Z$. Then from the table we have

$$P(1 \leq X) = P(-0.5 \leq Z) = P(Z \leq 0.5) = \Phi(0.5) = 0.6915.$$

(c) Since

$$P(X \leq 1) = 1 - P(1 \leq X) = 1 - 0.6915 = 0.3085.$$

We have

$$P(1 \leq X \leq 3) = P(X \leq 3) - P(X \leq 1) = 0.6915 - 0.3085 = 0.3830.$$

Exercise 3. Let x follows the normal distribution $N(1, 9)$. Find
(a) $P(X \leq 1.4)$; (b) $P(X \leq -1.22)$; (c) Hence find $P(-1.22 \leq X \leq 1.4)$.

Example 6. Find k such that $P(Z \geq k) = 0.1$ where z follows the standard normal distribution.

We note that for that k, we have

$$P(Z \leq k) = 1 - 0.1 = 0.9.$$

From the table we have

$$\Phi(1.28) = 0.8997 \quad \text{and} \quad \Phi(1.29) = 0.9015.$$

Therefore we know $k \in (1.28, 1.29)$ and we try to approximate k by using a **linear approximation (linear interpolation):**

$$\frac{0.9 - 0.8997}{k - 1.28} = \frac{0.9015 - 0.8997}{1.29 - 1.28} = \frac{0.0018}{0.01} = 0.18.$$

Thus

$$k \approx 1.28 + \frac{1}{600}.$$

Exercise 4. Find k such that $P(-k \leq Z \leq k) = 0.97$ where Z is the standard normal random variable.

Example 7. Find $P(z \geq -1.285)$ where $z \sim N(0, 1)$.

• By symmetry of the standard normal distribution, we have

$$P(z \geq -1.285) = P(z \leq 1.285) = q$$

• However, 1.285 has three decimal places and therefore we cannot find the probability in the table.

• But we have

$$P(z \leq 1.29) = 0.9015 \quad \text{and} \quad P(z \leq 1.28) = 0.8997$$

and therefore $0.8897 < q < 0.9015$. We then apply the **linear approximation method** again:

$$\frac{q - 0.8997}{1.285 - 1.28} = \frac{0.9015 - 0.8997}{1.29 - 1.28}$$

and

$$q = 0.8997 + 0.18 \times 0.005 = 0.9006.$$

15.4 Some Special Probability

- We have

$$P(-1 \le Z \le 1) \approx 0.68$$

and

$$P(-2 \le Z \le 2) \approx 0.95$$

and

$$P(-3 \le Z \le 3) \approx 0.99.$$

- In order words for any x follows $N(\mu, \sigma^2)$ we have

$$P(\mu - \sigma \le X \le \mu + \sigma) \approx 0.68$$

and

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.95$$

and

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.99.$$

16 Applications of Normal Distribution

Example 8. Suppose the salary of a group of 1000 civil servants follows the normal distribution $N(10000, 1000^2)$.
(a) Find the number of civil servants having salary less than 10500.
(b) What is the lowest salary of the top 200 civil servants?

We note that $\mu = 10000$ and $\sigma = 1000$.

(a) Let X be the salary of a civil servant, then we have

$$p = P(X \leq 10500) = P\left(Z \leq \frac{(10500 - 10000)}{1000}\right) = \underbrace{P(Z \leq 0.5) = 0.6915}_{\text{from the N(0,1) table}}.$$

Here the probability is obtained from the table.

Thus the number of civil servants is

$$1000 \times 0.6915 \approx 692.$$

(b) We are to find k such that

$$P(X \leq k) = \frac{1000 - 200}{1000} = 0.8$$

i.e.,

$$P\left(Z \leq \frac{k - 10000}{1000}\right) = \Phi\left(\frac{k - 10000}{1000}\right) = \Phi(K) = 0.8.$$

We note that

$$\Phi(0.84) = 0.7995 \quad \text{and} \quad \Phi(0.85) = 0.8023.$$

Thus $0.84 \leq K \leq 0.85$, and we apply linear approximation and get

$$\frac{0.8 - 0.7995}{K - 0.84} = \frac{0.8023 - 0.7995}{0.85 - 0.84}.$$

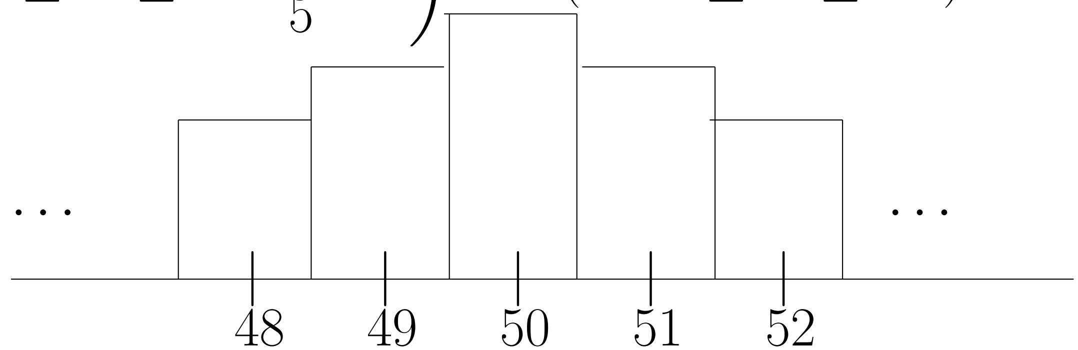
Therefore $K = 0.8418 = (k - 10000)/1000$ and hence $k = 10842$.

Exercise 5. A machine produces tubes of length 1m. Assume the length of the tubes follows the normal distribution $N(1, 0.04)$. If the length of a tube has a deviation less than 0.1m from the mean (Class A), the profit is 100 dollars. If the deviation more than 0.1m but less than 0.2 from the mean (Class B), the profit is 50 dollars. However, if the deviation is more than 0.2 (Class C), then it incurs a loss of 80 dollars. If 1000 tubes are produced, what will be expected profit?

For a large λ, $Poi(\lambda)$ can be approximated by $N(\lambda, \lambda)$. And for a large n, a Binomial r.v. follows $Bin(n, p)$ can be approximated by $N(np, np(1-p))$.

Example 9. A fair coin is tossed 100 times. Find the probability that the number of heads obtained is between 48 and 52 by using the normal approximation. We note that $\mu = 100 \times 0.5 = 50$ and $\sigma^2 = np(1-p) = 100 \times 0.5 \times 0.5 = 25$. The area of the following five rectangles are approximated by $P(47.5 \leq X \leq 52.5)$.

$$P\left(\frac{47.5-50}{5} \leq Z \leq \frac{52.5-50}{5}\right) = P(-0.5 \leq Z \leq 0.5) = 1-2(1-0.6915) = 0.383.$$ [6]



Exercise 6. A biased coin is tossed 200 times (probability of getting a head is 0.5). Find the probability that the number of heads obtained is between 98 and 102 by using the normal distribution approximation.

[6] Since we are approximating a p.d.f. of a discrete random variable by a continuous one, some adjustment has to be made so as to get a better approximation.

16.2 Central Limit Theorem*

This section is optional and it aims at introducing the **Central Limit Theorem**.

Proposition 6. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, identically distributed random variables with mean μ and variance σ^2. Then the following random variable tends to the normal distribution with mean 0 and variance 1 as $n \to \infty$:*

$$\boxed{Z_n = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.}$$

This is called the **Central Limit Theorem**.

Example 10. Let us give a heuristic argument for the Stirling formula.

Let $Y_1, Y_2, \ldots, Y_n$ be n independent Poisson random variables having same mean 1 (i.e., they follow $Poi(1)$). Then

$$Z_n = Y_1 + Y_2 + \ldots + Y_n,$$

the sum of the n Poisson random variables is also a Poisson random variable with **mean** n and **variance** n (i.e., Z_n follows $Poi(n)$).

We have

$$\begin{aligned} P(Z_n = n) &= P(n - \tfrac{1}{2} \leq Z_n \leq n + \tfrac{1}{2}) \\ &= P(\frac{-1}{2\sqrt{n}} \leq \frac{Z_n - n}{\sqrt{n}} \leq \frac{1}{2\sqrt{n}}) \\ &\approx \int_{\frac{-1}{2\sqrt{n}}}^{\frac{1}{2\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \approx \int_{\frac{-1}{2\sqrt{n}}}^{\frac{1}{2\sqrt{n}}} \frac{1}{\sqrt{2\pi}} dx = \frac{1}{\sqrt{2\pi n}}. \end{aligned}$$

Because for large n we have

$$\boxed{\frac{Z_n - n}{\sqrt{n}} \sim N(0,1)}$$

and $e^{\frac{-x^2}{2}} \approx 1$ for $x \in (\frac{-1}{2\sqrt{n}}, \frac{1}{2\sqrt{n}})$.

Now since Z_n is a Poisson random variable

$$P(Z_n = n) = \frac{e^{-n} n^n}{n!}$$

therefore we have

$$n! \approx n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}.$$

A Summary

1. Bernoulli experiment
2. Bernoulli distribution $B(p)$
3. Geometric distribution $Geo(p)$
4. Binomial distribution $Bin(n, p)$
5. Negative Binomial distribution $B_r(k, p)$
6. Poisson distribution $Poi(\lambda)$
7. Normal distribution $N(\mu, \sigma^2)$
8. The standard normal distribution $N(0, 1)$
9. Linear (interpolation) approximation method
10. Normal approximates Binomial
11. Normal approximates Poisson