

## Binomial Distribution

$$P(X=x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

where value of  $X$  denotes the number of successes in  $n$  trials.

## Poisson distribution

Let  $X$  be a discrete random variable taking on the values  $0, 1, 2, \dots$  such that the probability function of  $X$  is given by:

$$f(x) = P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x=0, 1, 2, \dots$$

$$E[X] = \sum x P(x) = \sum_{k \geq 0} k \frac{1}{k!} \lambda^k e^{-\lambda}$$

$$= \lambda e^{-\lambda} \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1}$$

$$= \lambda e^{-\lambda} \sum_{j \geq 0} \frac{\lambda^j}{j!} \quad \text{by letting } j = k-1$$

But  $\sum_{j \geq 0} \frac{\lambda^j}{j!} = e^\lambda$  (power series expansion)

$$\Rightarrow E[X] = \lambda e^{-\lambda} e^\lambda = \lambda \Rightarrow \mu_X = \lambda$$

Similarly, one can show that  $\sigma_X^2 = \lambda$  so

$$\text{that } \sigma_X = \sqrt{\lambda}.$$

If  $n$  is large ( $n \geq 50$ ) while  $np < 5$ , the Binomial distribution is closely approximated by the Poisson distribution. "for rare events"

Ex) If the probability that an individual will suffer a bad reaction from injection of a flu vaccine is 0.001. Find the probability that from  $n=2000$  individuals, (a) exactly 3 and (b) more than 2 individuals will suffer a bad reaction.

(a) use binomial:  $P(X=3) = \binom{2000}{3} (0.001)^3 (1-0.001)^{1997}$

Use Poisson approximation (ok since  $n$  is large and  $p$  small):

$$\lambda = np = 2000(0.001) = 2$$

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} \Rightarrow P(X=3) = \frac{2^3 e^{-2}}{3!} \approx 0.180$$

$$\begin{aligned}
 P(x > 2) &= 1 - [P(x=0) + P(x=1) + P(x=2)] = 1 - P(x \leq 2) \\
 &= 1 - \left[ \frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!} \right] \\
 &= 1 - 5e^{-2} \approx 0.323
 \end{aligned}$$

## Normal Distribution

density function given by:

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

$$P(x \leq x) = \int_{-\infty}^x f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv$$

$z = \frac{x-\mu}{\sigma}$  is std normal (mean zero,  $\sigma_z = 1$ )

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{symmetric wrt to } z=0$$

$$P(z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^z e^{-u^2/2} du$$



area  
under left  
half

given by  
table

If  $n$  is large and if neither  $p$  nor  $q$  is too close to zero, the binomial distribution can be closely approximated by normal distribution with std normal r.v.:

$$z = \frac{x - np}{\sqrt{npq}}$$

For  $X$  binomial r.v.:

$$E[X] = \mu_x = np$$

$$\sigma_x = \sqrt{npq}$$

$$np > 5$$

$$nq > 5$$

Ex) A fair coin is tossed 500 times. Find the probability that the number of heads will not differ from 250 by more than 10.

$$P(240 \leq \underbrace{X_n}_{\text{binomial}} \leq 260) \approx P(\underbrace{239.5 \leq X_n \leq 260.5}_{\text{normal with continuity correction}})$$

$$= P\left(\frac{239.5 - 250}{11.18} \leq Z_n \leq \frac{260.5 - 250}{11.8}\right)$$

where  $\mu = np = (500)\left(\frac{1}{2}\right) = 250$  and

$$\sigma = \sqrt{npq} = \sqrt{500\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} \approx 11.18$$



$$= P(-0.94 \leq Z \leq 0.94) = 2P(0 \leq Z \leq 0.94)$$

$$= 2A(0.94) = 2(0.3264) = 0.6528$$

## Central Limit Theorem

Let  $X_1, X_2, \dots, X_n$  be independent random variables from some probability distribution with mean  $\mu$  and variance  $\sigma^2$ . Then if

$$S_n = X_1 + X_2 + \dots + X_n \quad \text{then: } E[S_n] = E[X_1] + \dots + E[X_n] \\ = \mu + \dots + \mu = n\mu$$

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2} du$$

For sample means:

if  $n$  large ( $n > 30$ ) then distribution of sample means  $\{\bar{X}\}$  for all possible samples of size  $n$  is approximately normal with mean  $\mu_{\bar{X}} = \mu$  and  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ .

What this says:  $E[\bar{X}] = \mu$  and  $\sigma_{\bar{X}}$  decreases with sample size. So values of sample mean cluster more and more closely around population mean as  $n$  increases.

confidence interval the probability  $(1-\alpha)$  that the interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.  $\alpha \in (0,1)$

An interval estimate of  $\theta$  is an interval of the form  $\hat{\theta}_1 < \theta < \hat{\theta}_2$  where  $\hat{\theta}_1$  and  $\hat{\theta}_2$  are appropriate values of r.v.  $\theta$  s.t:

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha \text{ for } \alpha \in (0,1).$$

Note: like point estimates, interval estimates of a given parameter are not unique.

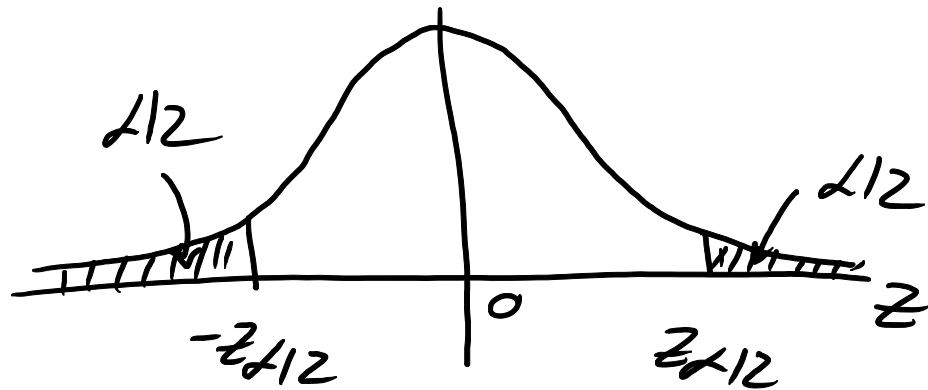
The  $z_{\alpha/2}$  value

Place area  $\alpha/2$  in each tail.

If we place area  $\alpha/2$  in each tail and if  $z_{\alpha/2}$  is the value of  $z$  such that area  $\alpha/2$  will lie to its right, then the confidence

interval with confidence coefficient  $(1-\alpha)$  (the probability that an interval estimator encloses a population parameter) is

$$\bar{X} \pm z_{\alpha/2} \sigma_{\bar{X}}$$



$$P(|z| < z_{\alpha/2}) = 1 - \alpha$$

where  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

$$\Rightarrow P(|\bar{x} - \mu_{\bar{x}}| < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$\Rightarrow$  Thm] If  $\bar{X}$ , the mean of a random sample of size  $n$  from a normal population with known variance  $\sigma^2$  is to be used as an estimator of the mean of the population,

the probability is  $(1-\alpha)$  that the error  
will be less than  $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ .