0.1 Normal Approximation to the Binomial

- Poisson distribution can be used to approximate binomial probabilities when n is quite large and p is very close to 0 or 1.
- Normal distribution not only provide a very accurate approximation to binomial distribution when n is large and p is not extremely close to 0 or 1,
- But also provides a fairly good approximation even when n is small and p is reasonably close to $\frac{1}{2}$.

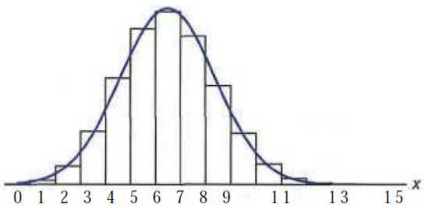

Figure 1: Normal approximation of $b(x; 15, 0.4)$.

- **Theorem 6.2**:
  If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the limiting form of the distribution of
  $$Z = \frac{X - np}{\sqrt{npq}}$$
  as $n \to \infty$, is the standard normal distribution $n(z; 0, 1)$
- $P(7 \leq X \leq 9)$

$$\sum_{x=7}^{9} b(x; 15, 0.4) = \sum_{x=0}^{9} b(x; 15, 0.4) - \sum_{x=0}^{6} b(x; 15, 0.4)$$
$$= 0.9662 - 0.6098 = 0.3564$$
$$\mu = np = 15 * 0.4 = 6, \ \sigma^2 = 15 * 0.4 * 0.6 = 3.6, \ \sigma = 1.897$$
$$z_1 = \frac{6.5 - 6}{1.897} = 0.26, \ and \ z_2 = \frac{9.5 - 6}{1.897} = 1.85$$
$$P(7 \leq X \leq 9) \approx P(0.26 < Z < 1.85) = P(Z < 1.85) - P(Z < 0.26)$$
$$= 0.9687 - 0.6026 = 0.3652$$

$P(X = 4) = b(4; 15, 0.4) = 0.1268$

$$z_1 = \frac{3.5 - 6}{1.897} = -1.32, \ and \ z_2 = \frac{4.5 - 6}{1.897} = -0.79$$

$$P(X = 4) \approx P(3.5 < X < 4.5) = P(-1.32 < Z < -0.79)$$
$$= P(Z < -0.79) - P(Z < -1.32)$$

$$= 0.2148 - 0.0934 = 0.1214$$

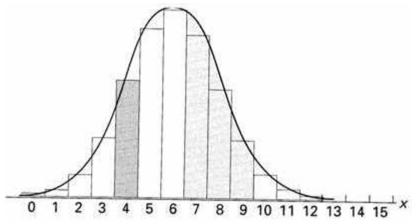

Figure 2: Normal approximation of $b(x; 15, 0.4)$ and $\sum_{x=7}^{9} b(x; 15, 0.4)$.

- The degree of accuracy, which depends on how well the curve fits the histogram, will increase as n increases.
- If both np and nq are greater than or equal to 5, the normal approximation will be good.

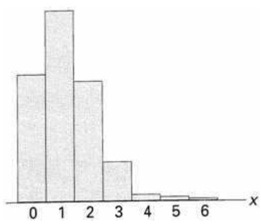

Figure 3: Histogram for $b(x; 6, 0.2)$.

- Let X be a binomial random variable with parameters n and p.

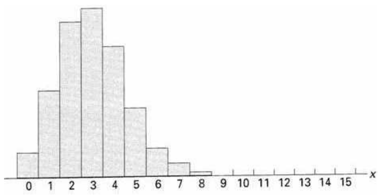

Figure 4: Histogram for $b(x; 15, 0.2)$.

- Then X has approximately a normal distribution with mean $\mu = np$ and variance $\sigma^2 = npq$ and

$$P(X \leq x) = \sum_{k=0}^{x} b(k; n, p)$$

$$\approx \textit{area under normal curve to the left of } x + 0.5$$

$$= P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{npq}}\right)$$

  and the approximation will be good if np and nq are greater than or equal to 5.

- **Example 6.15**: The probability that a patient recovers from a rare blood disease is 0.4.

- If 100 people are known to have contracted this disease, what is the probability that less than 30 survive?

- Solution:

$$\mu = np = 100 * 0.4 = 40$$
$$\sigma = \sqrt{100 * 0.4 * 0.6} = 4.899$$
$$z_1 = \frac{29.5 - 40}{4.899} = -2.14$$
$$P(X < 30) \approx P(Z < -2.14)$$
$$= 0.0162$$

- **Example 6.16**: A multiple-choice quiz has 200 questions each with 4 possible answers of which only 1 is correct answer.

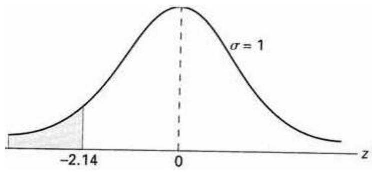

Figure 5: Area for Example 6.15.

- What is the probability that sheer guess-work yields from 25 to 30 correct answers for 80 of the 200 problems about which the student has no knowledge?
- Solution:

$$\mu = np = 80 * \frac{1}{4} = 20$$
$$\sigma = \sqrt{80 * \frac{1}{4} * \frac{3}{4}} = 3.873$$
$$z_1 = \frac{24.5 - 20}{3.873} = 1.16,$$
$$z_2 = \frac{30.5 - 20}{3.873} = 2.71$$
$$P(25 \leq X \leq 30) = \sum_{x=25}^{30} b(x; 80, \frac{1}{4})$$
$$\approx P(1.16 < Z < 2.71)$$
$$= 0.9966 - 0.8770 = 0.1196$$

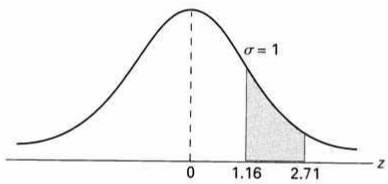

Figure 6: Area for Example 6.16.

0.2 Gamma and Exponential Distributions

- Exponential is a special case of the gamma distribution.
- Play an important role in queuing theory and reliability problems.
- Time between arrivals at service facilities, time to failure of component parts and electrical systems.
- **Definition 6.2**:

The **gamma function** is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx,\ for\ \alpha > 0$$

with

$$\Gamma(n) = (n-1)(n-2)\ldots\Gamma(1),$$

$$\Gamma(n) = (n-1)!\ with\ \Gamma(1) = 0! = 1,$$

- also

$$\Gamma(n+1) = n\Gamma(n) = n!$$

$$\Gamma(1/2) = \sqrt{\pi}\ exception$$

- **Gamma Distribution**: The continuous random variable X has a gamma distribution, with parameters α and β,
- If its density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}, & x > 0 \\ 0, & elsewhere \end{cases}$$

  where $\alpha > 0$ and $\beta > 0$

- The mean and variance of the gamma distribution are (Proof is in Appendix A.28)

$$\mu = \alpha\beta\ and\ \sigma^2 = \alpha\beta^2$$

- **Exponential Distribution** ($\alpha = 1$, special gamma distribution): The continuous random variable X has an exponential distribution, with parameters β,
- In real life, we observe the lifetime of certain products decreased as time goes.

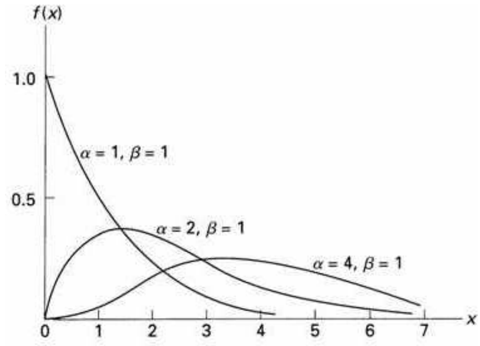

Figure 7: Gamma Distributions.

- To model life-lengths, especially the exponential curve seemed be good to fit these data rather well.
- If its density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0, & elsewhere \end{cases}$$

  where $\beta > 0$

- The mean and variance of the exponential distribution are

$$\mu = \beta \ and \ \sigma^2 = \beta^2$$

- The exponential distribution has a single tail. The single parameter β determines the shape of the distribution.
- **Relationship to the Poisson Process**: The most important applications of the exponential distribution are situations where the Poisson process applies.
- An industrial engineer may be interested in modeling the time T between arrivals at a congested interaction during rush hour in a large city. An arrival represents the **Poisson event**.
- Using Poisson distribution, the probability of no events occurring in the span up to time t

$$p(0, \lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

- Let X be the time to the first Poisson event.
- The probability that the length of time until the first event will exceed x is the same as the probability that no Poisson events will occur in x.

$$P(X \geq x) = e^{-\lambda x} \Rightarrow P(0 \leq X \leq x) = 1 - e^{-\lambda x}$$

- Differentiate the cumulative distribution function for the exponential distribution

$$f(x) = \lambda e^{-\lambda x} \; with \; \lambda = 1/\beta$$

- **Applications of Gamma and Exponential Distributions**
- The mean of the exponential distribution is the parameter β, the reciprocal of the parameter in the Poisson distribution.
- Poisson distribution has no memory, implying that occurrences in successive time periods are independent. They immediately "forget" their past behavior.
- The important parameter β is the mean time between events.
- The equipment failure often conforms to this Poisson process, β is called mean time between failures.
- Many equipment breakdowns do follow the Poisson process, and thus the exponential distribution does apply.
- Other applications include survival times in bio-medical experiments and computer response time.
- **Example 6.17**: Suppose that a system contains a certain type of component whose time in years to failure is given by T.
- The random variable T is modeled nicely by the exponential distribution with mean time to failure $\beta = 5$.
- Solution:

$$P(T > 8) = \frac{1}{5} \int_8^\infty e^{-t/5} dt = e^{-8/5} \approx 0.2$$

Let X represent the number of components functioning after 8 years.

- If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years.
- Solution:

$$P(X \geq 2) = \sum_{x=2}^{5} b(x; 5, 0.2) = 1 - \sum_{x=0}^{1} b(x; 5, 0.2)$$
$$= 1 - 0.7373 = 0.2627$$

- **Example 6.18**: Suppose that telephone calls arriving at a switchboard follow a Poisson process with an average of 5 calls coming per minute.
- What is the probability that up to a minute will elapse until 2 calls have come in to the switchboard?
- Solution:

  The Poisson process applies with time until 2 Poisson events following a gamma distribution with $\beta = 1/5$ and $\alpha = 2$.
  Let represent the time in minutes that transpires before 2 calls come.

$$P(X \leq x) = \int_0^x \frac{1}{\beta^2} x e^{-x/\beta} dx$$

$$P(X \leq 1) = 25 \int_0^1 x e^{-5x} dx$$
$$= 1 - e^{-5*1}(1 + 5) = 0.96$$

- **Example 6.19**: In a biomedical study with rats a dose-response investigation is used to determine the effect of the dose of a toxicant on their survival time.
- For a certain dose of the toxicant the study determines that the survival time, in weeks, has a gamma distribution with $\alpha = 5$ and $\beta = 10$.
- What is the probability that a rat survives no longer than 60 weeks?
- Solution:

  Let X be the survival time

$$P(X \leq x) = \int_0^x \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx$$

TABLE A.24 The Incomplete Gamma Function: $F(x;\alpha) = \int_0^x \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy$

x \ α	1	2	3	4	5	6	7	8	9	10
1	0.632	0.264	0.080	0.019	0.004	0.001	0.000	0.000	0.000	0.000
2	0.865	0.594	0.323	0.143	0.053	0.017	0.005	0.001	0.000	0.000
3	0.950	0.801	0.577	0.353	0.185	0.084	0.034	0.012	0.004	0.001
4	0.982	0.908	0.762	0.567	0.371	0.215	0.111	0.051	0.021	0.008
5	0.993	0.960	0.875	0.735	0.560	0.384	0.238	0.133	0.068	0.032
6	0.998	0.983	0.938	0.849	0.715	0.554	0.394	0.256	0.153	0.084
7	0.999	0.993	0.970	0.918	0.827	0.699	0.550	0.401	0.271	0.170
8	1.000	0.997	0.986	0.958	0.900	0.809	0.687	0.547	0.407	0.283
9		0.999	0.994	0.979	0.945	0.884	0.793	0.676	0.544	0.413
10		1.000	0.997	0.990	0.971	0.933	0.870	0.780	0.667	0.542
11			0.999	0.995	0.985	0.962	0.921	0.857	0.768	0.659
12			1.000	0.998	0.992	0.980	0.954	0.911	0.845	0.758
13				0.999	0.996	0.989	0.974	0.946	0.900	0.834
14				1.000	0.998	0.994	0.986	0.968	0.938	0.891
15					0.999	0.997	0.992	0.982	0.963	0.930

$$P(X \leq 60) = \frac{1}{\beta^5} \int_0^{60} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(5)} dx$$

Using incomplete gamma function $F(x;\alpha) = \int_0^x \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)} dy$
Let $y = x/\beta$, and $x = \beta y$

$$\Rightarrow P(X \leq 60) = \int_0^6 \frac{y^4 e^{-y}}{\Gamma(5)} dy$$
$$= F(6;5) = 0.715, \ \textit{see Appendix A.24}$$

0.3 Chi-Squared Distribution

- **Chi-Squared Distribution** ($\alpha = \nu/2$ and $\beta = 2$, special gamma distribution): The continuous random variable X has a chi-squared distribution, with ν degrees of freedom, if its density function is given by
$$f(x;\nu) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, & x > 0 \\ 0, & elsewhere \end{cases}$$
where ν is a positive integer
- The chi-squared distribution plays a vital role in statistical inference.

- Topics dealing with sampling distributions, analysis of variance and nonparametric statistics involve extensive use of the chi-squared distribution.

- **Theorem 6.4**:

  The mean and variance of the chi-squared distribution are

  $$\mu = \nu \textit{ and } \sigma^2 = 2\nu$$

0.4 Lognormal Distribution

- The lognormal distribution applies in cases where a natural log transformation results in a normal distribution.

- **Lognormal Distribution**: The continuous random variable X has a lognormal distribution if the random variable $Y = ln(X)$ has a normal distribution with mean μ and standard deviation σ.

- The resulting density function of X is

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}[(ln(x)-\mu)/\sigma]^2}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- The normal distribution has 2-tails. The lognormal distribution has a single tail.

- The normal distribution extends to $-\infty$ and ∞.

- The lognormal only extends to ∞ but is 0 for $x < 0$.

- **Theorem 6.5**:

  The mean and variance of the lognormal distribution are

  $$\mu = e^{\mu+\sigma^2/2} \textit{ and } \sigma^2 = e^{2\mu+\sigma^2} * (e^{\sigma^2} - 1)$$

- **Example 6.22**: Suppose it is assumed that the concentration of a certain pollutant produced by chemical plants, in parts per million, has a lognormal distribution with parameters $\mu = 3.2$ and $\sigma = 1$.

- What is the probability that the concentration exceeds 8 parts per million? (Table A.3)

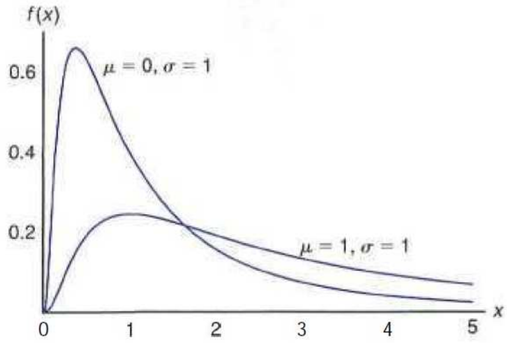

Figure 8: Lognormal Distributions.

- Solution:

  Let X be the pollutant concentration
  Since $ln(X)$ has a normal distribution with $\mu = 3.2$ and $\sigma = 1$

$$P(X > 8) = 1 - P(x \leq 8) = 1 - \Phi\left[\frac{ln(8) - 3.2}{1}\right]$$

$$= 1 - \Phi\left[\frac{2.08 - 3.2}{1}\right] = 1 - \Phi(-1.12)$$
$$= 1 - 0.1314 = 0.8686$$

  Here,we use the Φ notation to denote the cumulative distribution function of the standard normal distribution.

- **Example 6.23**: The life, in thousands of miles, of a certain type of electronic control for locomotives has an approximate lognormal distribution with $\mu = 5.149$ and $\sigma = 0.737$.

- Find the 5^{th} percentile of the life of such locomotive?

- Solution:

$$P(Z < z_1) = 0.05 \Rightarrow z_1 = -1.645$$

  $ln(x)$ has a normal distribution with $\mu = 5.149$ and $\sigma = 0.737$

$$\frac{ln(x) - 5.149}{0.737} = -1.645$$

$$\Rightarrow ln(x) = 0.737 * (-1.645) + 5.149 = 3.937$$
$$\Rightarrow x = 51.265$$

5% of the locomotives will have lifetime less than 51.265 thousand miles