Probability B: Univariate Distributions

Probability Density and Distribution for *Discrete* Random Variables

- **Definition:** Probability Distribution or Function (p 211). A ***table***, ***graph***, **or** ***mathematical function*** that describes all ***potential outcomes*** x of the random variable X from its ***underlying population*** and their corresponding probabilities.

- Notation: $\Pr(X = x)$ where x is an outcome of random variable X.

- Recall the properties of these probabilities following Kolmogorov's axioms.

  - [a] $0 \leq \Pr(X = x) \leq 1$

  - [b] $\Pr(a \leq X \leq b) = \sum_{x \in [a,b]} \Pr(X = x)$

  - [c] $\Pr(\sum_{X \in \mathcal{X}} \Pr(X = x)) = 1$ because the set of all ***elementary outcomes*** are ***mutually exclusive*** and ***exhaustively cover*** the sample space $\mathcal{X}$ (all possible outcomes) of the random variable X.

- Representation as graph or table:

TABLE 6-2
Relative Frequency Probabilities for the Random Variable Called Household Size of Table 6-1

x_i	$P(x_i)$ or $P(X = x_i)$
2	1/8 = .125
3	2/8 = .250
4	3/8 = .375
5	2/8 = .250
	1.0

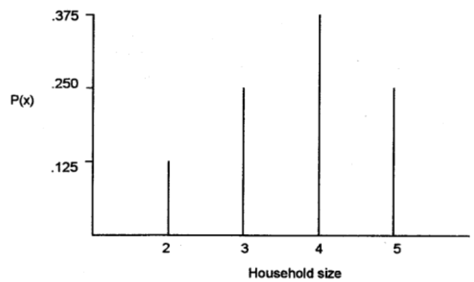


FIGURE 6-1. A probability mass function.

- If the individual data values x_i are sorted ascending, i.e., $x_1 < x_2 < \ldots < x_n$, then we can express the ***cumulative probability*** function, denoted by a capital $F(\,)$, as

$$F(x_i) = \sum_{x \le x_i} \Pr(X = x_i) = \Pr(X = x_1) + \Pr(X = x_2) + \cdots + \Pr(X = x_i)$$

- Note: we always can ***retrieve the individual probabilities*** from the cumulative probability function:

$$\Pr(X = x_1) = F(x_1)$$
$$\Pr(X = x_2) = F(x_2) - F(x_1)$$
$$\vdots$$
$$\Pr(X = x_n) = F(x_n) - F(x_{n-1})$$

TABLE 6-4
Cumulative Mass Function

x	$P(x_i)$	$F(x_i)$
2	0.125	0.125
3	0.250	0.375
4	0.375	0.750
5	0.250	1.000

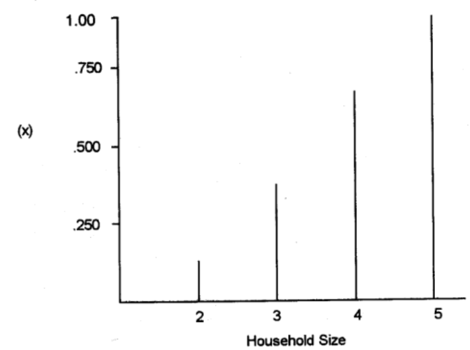


FIGURE 6-2. A cumulative probability mass function.

Expectation and Variance of Discrete Random Variables

- The ***expected value*** and the ***variance*** are population counterparts to the sample statistics of the estimated mean and estimated variance.
  Note that the summation has to proceed over the complete range of possible values of X.

- **Definition:** Expected Value. For a discrete random variable X with values $x_1, x_2, \ldots, x_k$, its expected value is

$$E(X) = \sum_{i=1}^{k} x_i \cdot \Pr(X = x_i)$$

- **Definition:** Variance. For a discrete random variable X with values $x_1, x_2, \ldots, x_k$, its variance is

$$\begin{aligned} Var(X) &= \sum_{i=1}^{k} \left[x_i - E(X)\right]^2 \cdot \Pr(X = x_i) \\ &= \sum_{i=1}^{k} x_i^2 \cdot \Pr(X = x_i) - \left[E(X)\right]^2 \end{aligned}$$

- In both expressions the constant sample weight $1/n$, that has been used when calculating the sample means and variances, has been replace by the probability $\Pr(X = x_i)$.

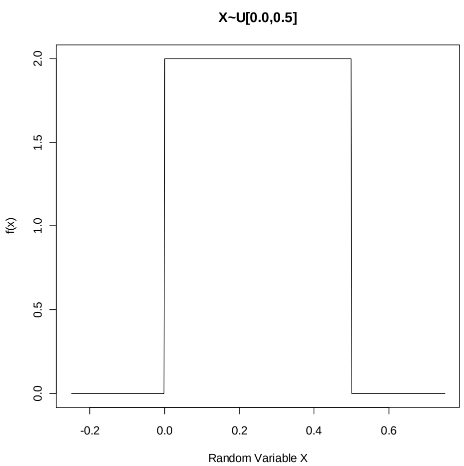
Probability Density and Distribution *Continuous* Random Variables

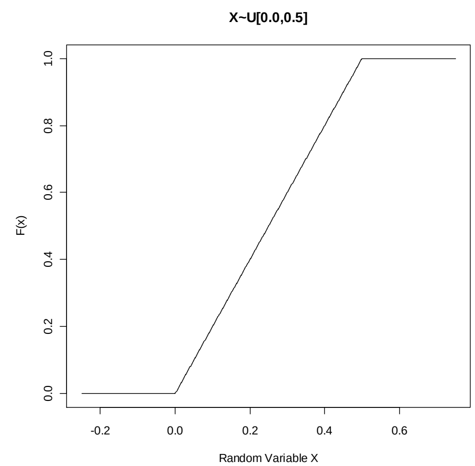
- Since we are dealing with an ***infinite number of representations*** of a random variable in its ***support*** along the ***real domain*** $[a,b]$ with the possible bounds $-\infty \le a$ and $b \le \infty$ it follows that:
  - the probability at an ***individual representation*** $X = x_i$ ***becomes zero***, i.e., $\Pr(X = x_i) = 0$.
    Otherwise, the probability over all representations may sum to a value larger than one.
  - The individual probabilities are replaced by the continuous ***probability density*** function $f(x)$ with
    $$f(x) = \begin{cases} > 0 & \text{for } x \in [a,b] \\ 0 & \text{for } x \notin [a,b] \end{cases}$$
    and the area under the density function has to ***integrate***[1] ***to one***: $\int_a^b f(x) \cdot dx = 1$.
  - Probabilities of subsets $[c,d]$ with $a \le b < c \le d$ can still be expressed by integrals:
    $$\Pr(c \le X \le d) = \int_c^d f(x) \cdot dx$$

[1] An integral is the area underneath the function $f(x)$ within the interval from a to b.

- Note: the density function at a given values $X = x$ can be larger than one, i.e., $f(x) > 1$.
  Example: $X \sim \mathrm{U}(0.0, 0.5)$ (read: the random variable X is distributed according to a uniform distribution with a support starting at 0 and ending at 0.5).
  Its density function is $f(x) = \begin{cases} 2 & x \in [0.0, 0.5] \\ 0 & \text{otherwise} \end{cases}$ with $\int_a^b f(x) \cdot dx = 1$.

- Its cumulative distribution function becomes $F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(x) \cdot dx$





$U(0.0, 0.5)$ with density function $f(x) = 2$ $\quad$ $U(0.0, 0.5)$ with distribution function $F(x)$

Specific Discrete Probability Distribution

Each distribution has a specific ***underlying model*** that has generated the probabilities of the events.

- **Uniform distribution:**
  - For k representations numbered $x \in \{1, 2, \ldots, k\}$ with equal probability the probability functions is

    $$\Pr(X = x \mid k) = \frac{1}{k} \text{ with } x \in \{1, 2, \ldots, k\}$$

  - The expectation is

    $$E(X) = \sum_{x=1}^{k} x \cdot \Pr(X = x \mid k) = \sum_{x=1}^{k} x \cdot \frac{1}{k}$$

    $$= \frac{1}{k} \cdot \left[\frac{k \cdot (k+1)}{2} \right] = \frac{k+1}{2}$$

  - The variance is $Var(X) = \dfrac{k^2 - 1}{12}$

- **Binomial distribution:**
  - Underlying assumptions:

    [a] There are n ***independent*** trials of an experiment, that is, the outcome of previous

trials will not influence the outcome of current or future trials.

[b] Only binary outcomes (e.g., success/failure, head/tail, male/female, 0/1, etc.) are possible at each trial.

[c] The probabilities π for successes and $1-\pi$ for failures remain constant for all trials.

[d] The random variable S is the total number of "successes" (sum of the individual successes) after n trials have been completed, that is, $0 \leq S \leq n$.

- Each individual trial follows a *Bernoulli* distribution (just one trial, i.e., $n = 1$)

$$\Pr(S = s) = \pi^s \cdot (1-\pi)^{1-s} \text{ with } s = \begin{cases} 1 & \text{for a success} \\ 0 & \text{for a failure} \end{cases}.$$

- Example: The joint probability of $\{S_1 = 0 \cap S_2 = 1\}$, that is one success $S = 1$ in two ***independent*** trails $n = 2$, becomes

$$\Pr(S_1 = 0 \cap S_2 = 1) = \left[\underbrace{\pi^0}_{=1} \cdot \underbrace{(1-\pi)^{1-0}}_{=1-\pi} \right] \cdot \left[\underbrace{\pi^1}_{=\pi} \cdot \underbrace{(1-\pi)^{1-1}}_{=1} \right]$$

This probability is identical for the individual event $\{S_1 = 1 \cap S_2 = 0\}$, therefore, order does not matter.

- The events $\{S_1 = 0 \cap S_2 = 1\}$ and $\{S_1 = 1 \cap S_2 = 0\}$ are both mutually exclusive (thus their probabilities can be summed up), thus

$$\begin{aligned}\Pr(X = 1 \mid n = 2, \pi) &= \Pr(S_1 = 0 \cap S_2 = 1) + \Pr(S_1 = 1 \cap S_2 = 0) \\ &= 2 \cdot \pi \cdot (1 - \pi)\end{aligned}$$

- The general equation for the binomial distribution takes the sum of successes from experiments with a difference sequence of successes into account

$$\Pr(X = x \mid n, \pi) = \frac{n!}{x! \cdot (n - x)!} \cdot \pi^x \cdot (1 - \pi)^{n-x}$$

- Example:

TABLE 6-7
Possible Outcomes of the Coin Toss Experiment

$n = 1$				T		H			
$n = 2$			TT		TH HT		HH		
$n = 3$		TTT		TTH THT HTT		THH HTH HHT		HHH	
$n = 4$	TTTT		TTTH TTHT THTT HTTT		TTHH THTH HTTH THHT HTHT HHTT		THHH HTHH HHTH HHHT		HHHH

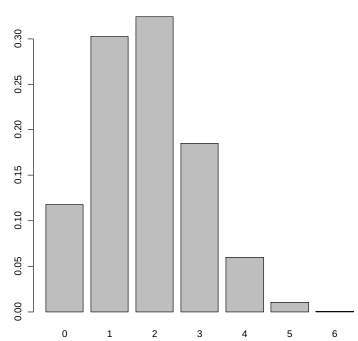
TABLE 6-9
Binomial Distributions

Number of trials n	Number of heads	Probability of number of heads
1	0	$1 - \pi$
	1	π
2	0	$(1 - \pi)^2$
	1	$2\pi(1 - \pi)$
	2	π^2
3	0	$(1 - \pi)^3$
	1	$3\pi(1 - \pi)^2$
	2	$3\pi^2(1 - \pi)$
	3	π^3
4	0	$(1 - \pi)^4$
	1	$4\pi(1 - \pi)^3$
	2	$6\pi^2(1 - \pi)^2$
	3	$4\pi^3(1 - \pi)$
	4	π^4

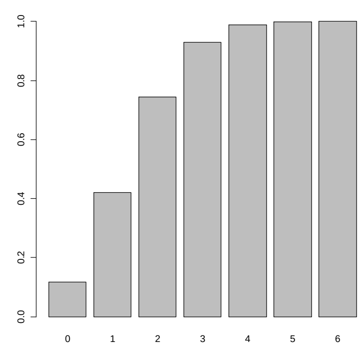
- The expectation of the binomial distribution is $E(X) = n \cdot \pi$ and the variance is $Var(X) = n \cdot \pi \cdot (1 - \pi)$

- Examples of the binominal distribution can be generated with the script **`BinomPoisson.r`**: $X \sim \text{Binomial}(\pi = 0.3, n = 6)$

Count	Pr(Count)
0	0.117649
1	0.302526
2	0.324135
3	0.185220
4	0.059535
5	0.010206
6	0.000729

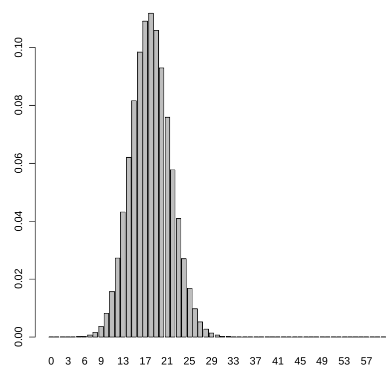


Binominal Density Function

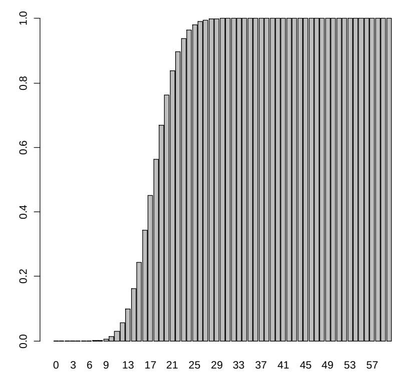


Binominal Distribution Function

- Examples of Binominal Distribution Functions: $X \sim \text{Binomial}(\pi = 0.3, n = 60)$



Binomial Density Function



Binomial Distribution Function

Note that for large *n* the discrete density almost looks like being continuous normal distributed

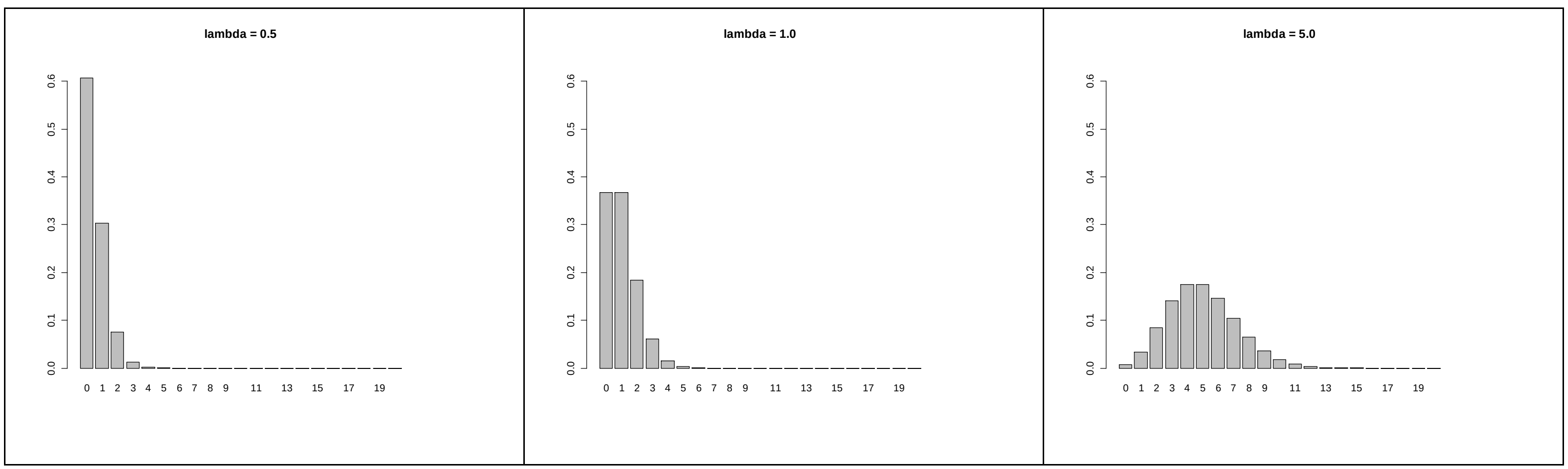
- **The Poisson Distribution (skipped)**
  - Important probability model for counting and queuing processes
  - Underlying assumptions:

    [a] The total number of events in two ***mutually exclusive*** intervals is independent

    [b] The probability of just one event in a small interval is small and proportional to the length of the interval, i.e., the event is rare

    [c] The probability of two or more events in a small interval is near zero

- The density of the Poisson distribution is $\Pr(X = x \mid \lambda) = \dfrac{\exp(-\lambda)\cdot\lambda^x}{x!}$ with $x \in \{0,1,2,\ldots\}$
- The expectation and variance are equal: $E(X) = Var(X) = \lambda$
- The Poisson distribution can be derived from the Binomial Distribution (see BBR, pp 228-230 and the R script **`BinomToPois.r`**) as the number of trials moves to infinity, i.e., $n \to \infty$, under the assumptions:
  [a] the number of intervals n per unit U increases the intervals become shorter U/n.
  [b] However, the probability of one event proportionally decreases $\Pr(X = 1) \sim U/n$.
  Note however, the expectation remains fixed, i.e., $E(X) = n \cdot \pi = \text{constant}$.
- Examples of Poisson Distributions for $\lambda \in \{0.5, 1.0, 5.0\}$:



- **Other frequently encountered discrete distributions (skipped)**
  - Geometric distribution: Number of Bernoulli trails before the first success occurs
  - Negative binominal distribution: Number of Bernoulli trials is $x+r$ until the x^{th} success occurs with r being the ***random number*** of failures. Therefore, the total number of experiments is not fixed.
    Can be viewed as a special case of the Binomial distribution where the number of trials is not fixed:
    $$\Pr(R=r|\pi,x)=\underbrace{\binom{r+x-1}{x-1}\cdot(1-\pi)^{r}\cdot\pi^{x-1}}_{r\text{ failures and }x-1\text{ successes}}\cdot\underbrace{\pi}_{x^{th}\text{ success}}.$$
    The minus one terms $r+x-1$ and $x-1$ appear because the order of outcomes for the first $r+x-1$ experiments is irrelevant. The last experiment must be a success and, therefore, the ***order of the last experiment matters***.
    The negative binominal distribution is also a generalization of the Poisson distribution where the variance is allowed to be larger than the expectation.
  - Multinomial distribution: More than two classes are allowed. The probability is associated with the number of observed counts in each class given fixed class probabilities.
    The Binomial distribution can be rewritten as

$$\Pr(X_1 = x_1, X_2 = x_2 | \pi_1, \pi_2; n) = \frac{n!}{x_1! \cdot x_2!} \cdot \pi_1^{x_1} \cdot \pi_2^{x_2}$$

under the constraints $\pi_1 + \pi_2 = 1$ and $x_1 + x_2 = n$. As long as these constraints are satisfied the probability for the counts $\{x_1, x_2, \dots, x_k\}$ in k classes with the class probabilities $\{\pi_1, \pi_2, \dots, \pi_k\}$ is given by

$$\Pr(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k | \pi_1, \pi_2, \dots \pi_k; n) = \frac{n!}{x_1! \cdot x_2! \cdots x_k!} \cdot \pi_1^{x_1} \cdot \pi_2^{x_2} \cdots \pi_k^{x_k}$$

- Hypergeometric distribution: Describes the distribution of number of samples x_1 obtained from one group with n_1 elements when sampling without replacement from two groups is considered. The second group has n_2 elements and the total sample size is $x_1 + x_2 = k$.

$$\Pr(X_1 = x_1, X_2 = x_2 | n_1, n_2; k) = \frac{\binom{n_1}{x_1} \cdot \binom{n_2}{x_2}}{\binom{n_1 + n_2}{x_1 + x_2}}$$

- **Wikipedia** provides good discussion of these discrete distributions.

Example: A surveying company has 8 identical GPS units. Three of these units received a software update, which later was identified to lead to a systematic error in the elevation estimate. For redundancy and cross-evaluation purposes, each surveying team is using two instruments in the field.

Calculate the probability that a surveying team did not spot the software error while doing measurements, that is, they were using two GPS units with faulty software updates?
Refer to the probability rules that you applied in your calculations and justify their use.

- Approach 1: $\Pr(S_1 = F,\ S_2 = F) = \frac{3}{8} \cdot \frac{2}{7} = \frac{6}{56} = 0.1071 \Rightarrow$ Sampling without replacement
- Approach 2: $\Pr(S_1 = F,\ S_2 = F) = \frac{|\text{Event Space}|}{|\text{Sample Space}|} = \frac{C_2^3 \cdot C_0^5}{C_2^8} = \frac{3 \cdot 1}{\frac{8 \cdot 7}{2}} = 0.1071 \Rightarrow$ Use of the hypergeometric rule and analytical definition of probabilities.

Specific continuous distributions

- **The Uniform distribution:**
  - It has two parameters describing its support. Its density is a constant density over its support.

$$f(x \mid a,b) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad F(x \mid a,b) = \begin{cases} \frac{x-a}{b-a} & \text{for } x \in [a,b] \\ 1 & \text{for } x > b \\ 0 & \text{for } x < a \end{cases}$$

- Its expectation is $E(X) = \dfrac{a+b}{2}$
- Its variance is $Var(X) = \dfrac{(b-a)^2}{12}$

- **The exponential distribution (skipped):**
  - It is usually viewed as the continuous analog to the Poisson distribution and expresses the probability of the wait time between events in a queue.
  - Its density function is

$$f(x \mid \lambda) = \begin{cases} \lambda \cdot \exp(-\lambda \cdot x) & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

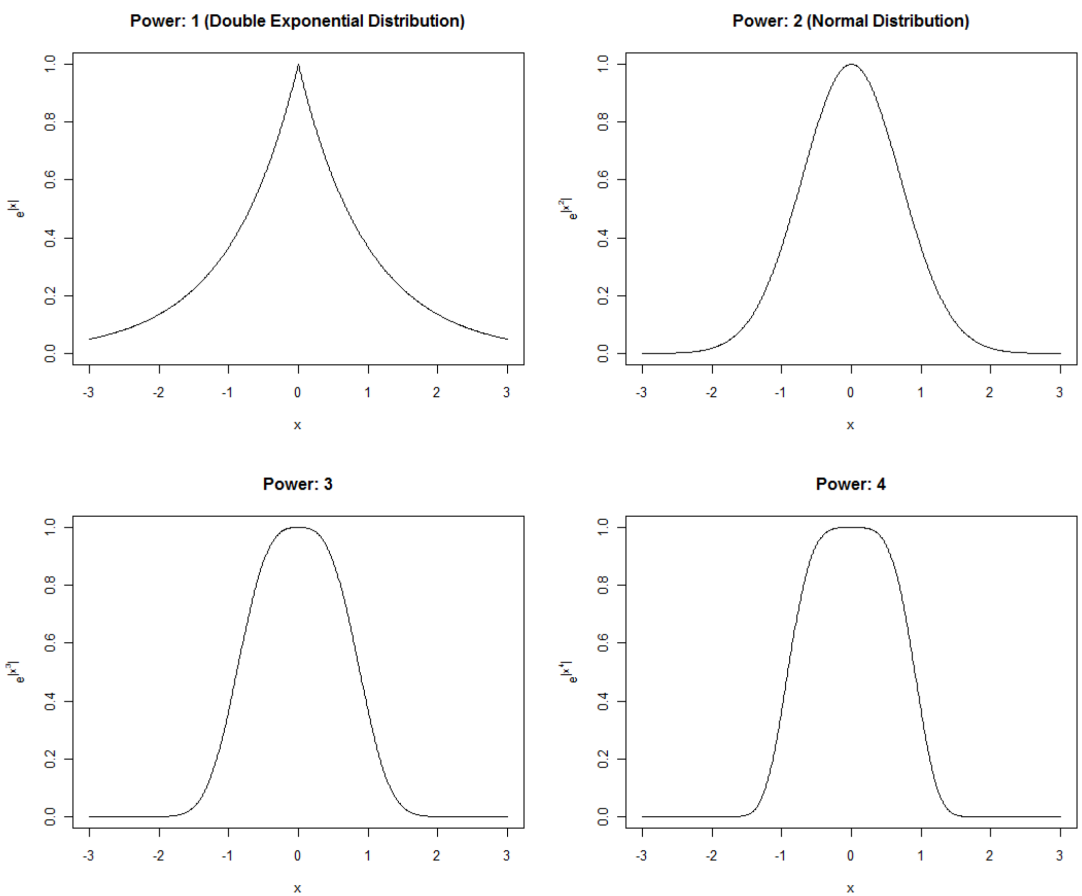
  - Its distribution function is

$$F(x \mid \lambda) = \begin{cases} 1 - \exp(-\lambda \cdot x) & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  - The expectation is $E(X) = 1/\lambda$ and the variance is $Var(X) = 1/\lambda^2$
  - The parameter λ can be estimated from sample observations by $\hat{\lambda} = 1/\bar{x}$.

- **The Normal (or Gaussian) distribution:**
  - One of the most important distribution:
    [a] It emerges in many natural processes such as the sum of many small random effects (=> more when we deal with the central limit theorem in the Chapter: Sampling)
    [b] Many other continuous distributions can be derived from the normal distribution (log-normal, *t*-, *F*-, and χ^2-distribution etc.)
    [c] Its multivariate equivalent has many elegant properties. E.g., marginal distributions again normal, the strength of a linear relationship between two random variables is related to their correlation coefficient etc.
    [d] Its conditional distribution remains a normal distribution.

- The shape of the *normal distribution* is based on an exponential function $f(x) \propto \exp(-|x|^p)$ with its argument p in the 2nd power :



- The univariate normal distribution $N(\mu,\sigma^2)$ is characterized by two parameters, μ is the central location parameter, σ^2 is its spread parameter:

  $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right]$ with $E(X) = \mu$ and $Var(X) = \sigma^2$

- Its cumulative distribution function cannot be given in a closed form but must be derived through numerical integration

  $F(x) = \int_{-\infty}^{x} \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right] \cdot dx$. Notice the upper bound x in the integral.

- The normal distribution is symmetric around its expected value, i.e., its skewness is zero and the mode, median and expectation are identical.
  Its kurtosis (balance between probability mass in the center and in the tails) is 0.

- **The z-transformation for Comparing the Shape of Distributions**

  - Any normal distributed variable $X \sim N(\mu, \sigma^2)$ can be expressed as a standard normal distributed variable $z(X) \sim N(0,1)$ using the transformation

$$z(X) = \frac{X - \mu}{\sigma} \ .$$

- Example of z-transformation:

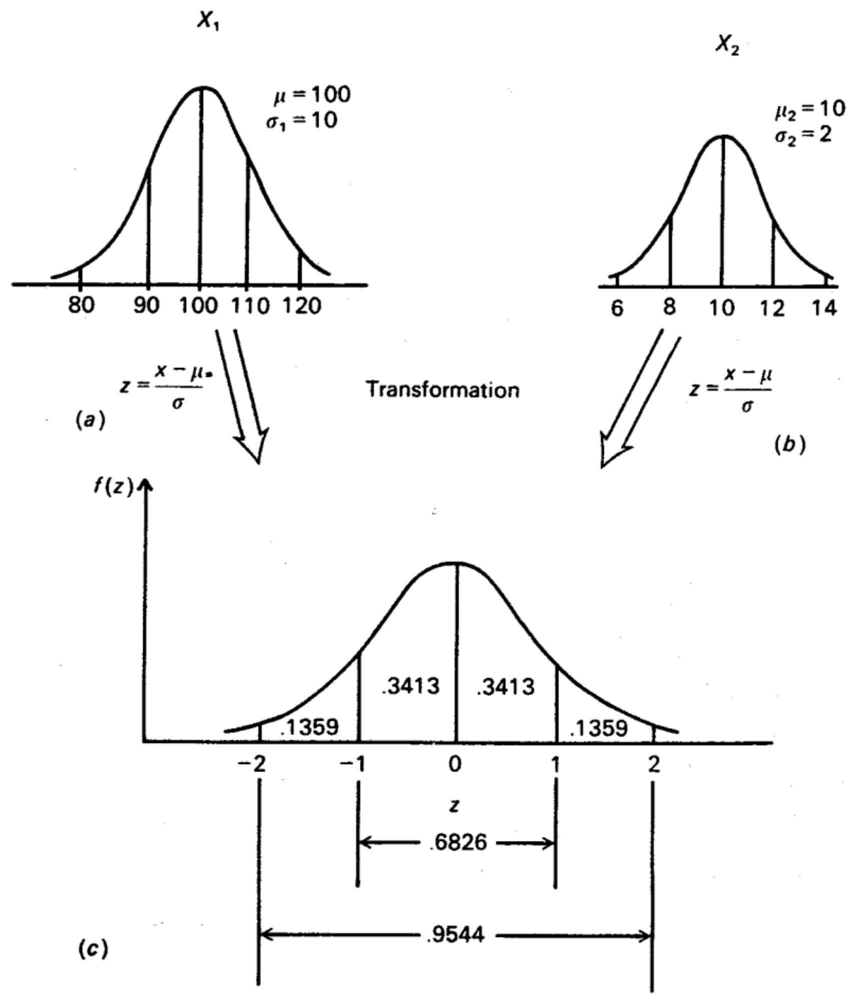


FIGURE 6-15. Converting normal distributions to the standard normal.

- Subtracting μ_X or $\bar{x}$ in the numerator shifts the distribution over the value 0 and dividing by the numerator σ_X or s_X rescales the distribution to a standard deviation of 1.

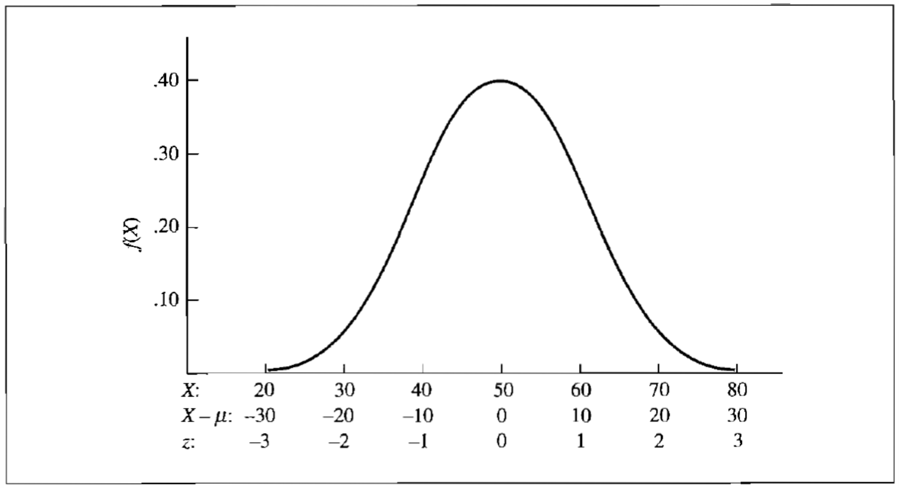


Figure 6.6
A normal distribution with various transformations on the abscissa

- The objective of the *z*-transformation is to transform a random variable X with the mean μ_X and variance σ_X^2 into a variable Z with mean $\mu_Z = 0$ and standard deviation $\sigma_Z^2 = \sigma_Z = 1$
  without changing the shape of the distribution.

- This transformation does not change the percentiles $p(x_{[i]}) = p(z_{[i]})$. Therefore, the distribution of metric variables measured in different units (e.g., Celsius versus Fahrenheit) become comparable.

The linear transformation

$$Z_i = \frac{X_i - \mu_X}{\sigma_X} \text{ for theoretical parameters or}$$

$$z_i = \frac{x_i - \bar{x}}{s_X} \text{ for estimated parameters}$$

achieves the desired task.

- Its reverse transformation back into the original units is

$$X_i = \sigma_X \cdot zZ_i + \mu_X \text{ or } x_i = s_X \cdot z_i + \bar{x}\text{, respectively}$$

- The standard function **`scale( )`** performs the z-transformation for all numeric vectors in a data-frame or columns of a numeric matrix.

- Several statistics can be rewritten in terms of the a z-transformed variable:

  - The skewness becomes $\text{skew}(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} = \sum_{i=1}^{n} z_i^3$

  - The kurtosis becomes $\text{kurt}(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3 = \sum_{i=1}^{n} z_i^4 - 3$

- The correlation coefficient becomes

$$r_{12} = \frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)\cdot(x_{i2}-\bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2\cdot\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} = \frac{\sum_{i=1}^{n} z_{i1}\cdot z_{i2}}{n-1}$$

- **Other important continuous distribution (skipped)**
  - Gamma distribution: the support is $0 \le X < \infty$. It is a generalization of exponential distribution
  - Beta distribution: the support $0 < X < 1$. It is used to model the distribution of probabilities and can take many different shapes depending on its underlying parameters, which can be estimated by the methods of moments.
  - Log-normal distribution: If $X \sim N(\mu_X, \sigma_X^2)$ then $Y = \exp(X)$ with $Y > 0$ follows a log-normal distribution with
    $E(Y) = \exp\left(\mu_X + \frac{1}{2}\cdot\sigma_X^2\right)$ and $Var(Y) = \exp(2\cdot\mu_X + \sigma_X^2)\cdot(\exp(\sigma_X^2) - 1)$.
    Analogy to the emergence of the normal distribution as a *sum* of many small random variables the log-normal distribution can be conceived of as a *product* of many small positive random effects.
  - χ^2 distribution: Let z_i be n independent standard normal distributed variables, then $\chi^2_{df=n} = \sum_{i=1}^{n} z_i^2$ is a χ^2 distributed random variable with n degrees of freedom.

- t distribution: Let z and z_i be $n+1$ independent standard normal distributed variables, then $t_{df=n} = \frac{z}{\sqrt{\frac{\sum_{i=1}^{n} z_i^2}{n}}}$ is a t distributed random variable with n degrees of freedom.

- F distribution: Let z_i and z_j be n and m distributed standard normal distributed variables, then $F_{df_1=n,df_2=m} = \frac{\frac{\sum_{i=1}^{n} z_i^2}{n}}{\frac{\sum_{j=1}^{m} z_j^2}{m}}$ is a F distributed random variable with n and m degrees of freedom.

- Again see Wikipedia for these and other continuous distributions.

Expectation and Variance of Continuous Distributions (skipped)

- Since the random variable is no longer countable in discrete increments, integration over the support has to replace summation.

- The probability of a random variable in an interval $X \in [a,b]$ with $a \le b$ is evaluated with the distribution functions $F(a) = \int_{-\infty}^{a} f(x) \cdot dx$ and $F(b) = \int_{-\infty}^{b} f(x) \cdot dx$ as $F(b) - F(a) = \int_{a}^{b} f(x) \cdot dx$

- Note the first derivate of the distribution function gives the density function:

$$\lim_{\Delta x \to 0} \frac{F(x) - F(x+\Delta x)}{\Delta x} = f(x)$$

- The expectation and variance require integration over the whole support of *X*:

  - Expectation: $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$

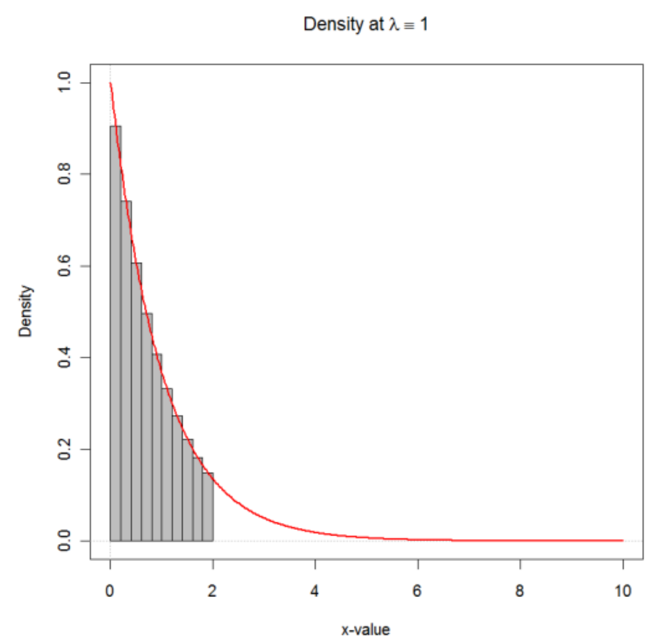
  - Variance: $$\begin{aligned} Var(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) \cdot dx \\ &= \int_{-\infty}^{\infty} x^2 \cdot f(x) \cdot dx - [E(X)]^2 \end{aligned}$$

- For some density functions their distribution function, expectation, and variance are known analytically. For others, however, they need to be approximated by numerical integration methods.

- Example: exponential distribution $f(x \mid \lambda = 1)$ with

  $F(2) = \int_0^2 f(x) \cdot dx = 0.8646647$

  The numerical approximated value using a Riemann sum with 10 bars is $F(2) \approx 0.8643045$.



Specific Distribution Functions

- Explore different distribution functions. Make sure to specify the lower or upper tails options properly:
    - **dunif** gives the density $f(x \mid a,b)$
    - **punif** gives the cumulative distribution $F(X < x \mid a,b)$
    - **qunif** gives the quantile x for a give probability q_x, that is, $F(X \leq x \mid a,b) = q_x$
    - **runif** draws random samples from $U(a,b)$
- Other short names for univariate distributions in are:
    - **[dpqr]pois** for the Poisson distribution
    - **[dpqr]binom** for the binomial distribution
    - **[dpqr]geom** for the geometric distribution
    - **[dpqr]nbinom** for the negative binomial distribution
    - **[dpqr]exp** for the exponential distribution
    - **[dpqr]beta** for the beta distribution
    - **[dpqr]gamma** for the gamma distribution
    - **[dpqr]norm** for the normal distribution

- **[dpqr]t** for the *t*-distribution
- **[dpqr]f** for the *F*-distribution
- **[dpqr]chisq** for the χ^2-distribution

Exercise with the Standard Normal Table (skipped)

TABLE A-3
Standard Normal Probabilities

z	$P(-z < Z < z)$	$P(\|Z\| > z)$	$P(Z > z)$	$P(Z < -z)$	$P(Z < z)$	$P(Z > -z)$
0.50	0.383	0.617	0.309	0.309	0.691	0.691
0.60	0.451	0.549	0.274	0.274	0.726	0.726
0.70	0.516	0.484	0.242	0.242	0.760	0.758
0.80	0.576	0.424	0.212	0.212	0.788	0.788
0.90	0.632	0.368	0.184	0.184	0.816	0.816
1.00	0.683	0.317	0.159	0.159	0.841	0.841
1.28	**0.800**	**0.200**	**0.100**	**0.100**	**0.900**	**0.900**
1.50	0.866	0.134	0.067	0.067	0.933	0.933
1.60	0.890	0.110	0.055	0.055	0.945	0.945
1.65	**0.900**	**0.100**	**0.050**	**0.050**	**0.950**	**0.950**
1.70	0.911	0.089	0.045	0.045	0.955	0.955
1.80	0.928	0.072	0.036	0.036	0.964	0.964
1.90	0.943	0.057	0.029	0.029	0.971	0.971
1.96	**0.950**	**0.050**	**0.025**	**0.025**	**0.975**	**0.975**
2.00	0.954	0.046	0.023	0.023	0.977	0.977
2.10	0.964	0.036	0.018	0.018	0.982	0.982
2.20	0.972	0.028	0.014	0.014	0.986	0.986
2.30	0.979	0.021	0.011	0.011	0.989	0.989
2.40	0.984	0.016	0.008	0.008	0.992	0.992
2.50	0.988	0.012	0.006	0.006	0.994	0.994
2.58	**0.990**	**0.010**	**0.005**	**0.005**	**0.995**	**0.995**
2.60	0.991	0.009	0.005	0.005	0.995	0.995
2.70	0.993	0.007	0.003	0.003	0.997	0.997
2.80	0.995	0.005	0.003	0.003	0.997	0.997
2.90	0.996	0.004	0.002	0.002	0.998	0.998
3.00	0.997	0.003	0.001	0.001	0.999	0.999
3.10	0.998	0.002	0.001	0.001	0.999	0.999
3.20	0.999	0.001	0.001	0.001	0.999	0.999
3.30	0.999	0.001	0.000	0.000	1.000	1.000
3.40	0.999	0.001	0.000	0.000	1.000	1.000

Probabilities C: Bivariate Distributions (skipped)

All output and plots in this section were generated with **`BiVariatePlots.r`**.

Joint Probabilities of Nominal and Ordinal Scaled Pairs of Variables

Nominal and ordinal scaled pairs of variables can be jointly displayed in contingency tables. Assuming that the 100 households are the population, then the relative frequencies become the joint probabilities $\Pr(C = i \cap P = j)$ and the marginal probabilities become $\Pr(P = j) = \sum_{i=0}^{3} \Pr(C = i \cap P = j)$

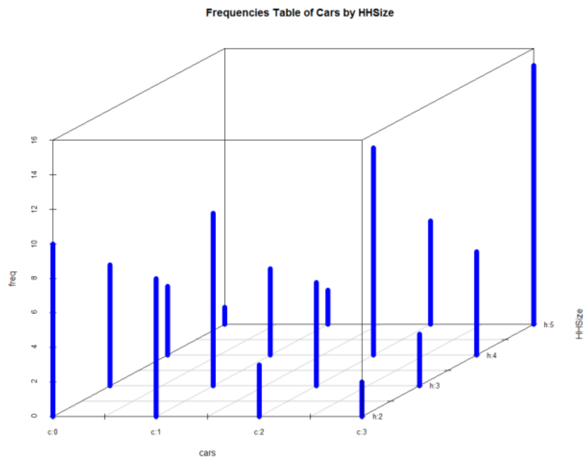
```
       HHSize
cars   p:2 p:3 p:4 p:5 Sum
  c:0   10   7   4   1  22
  c:1    8  10   5   2  25
  c:2    3   6  12   6  27
  c:3    2   3   6  15  26
  Sum   23  26  27  24 100
```

Frequency table n_{ij}

```
       HHSize
cars    p:2  p:3  p:4  p:5  Sum
  c:0 0.10 0.07 0.04 0.01 0.22
  c:1 0.08 0.10 0.05 0.02 0.25
  c:2 0.03 0.06 0.12 0.06 0.27
  c:3 0.02 0.03 0.06 0.15 0.26
  Sum 0.23 0.26 0.27 0.24 1.00
```

Relative frequency table $f_{ij} = n_{ij}/n_{++}$

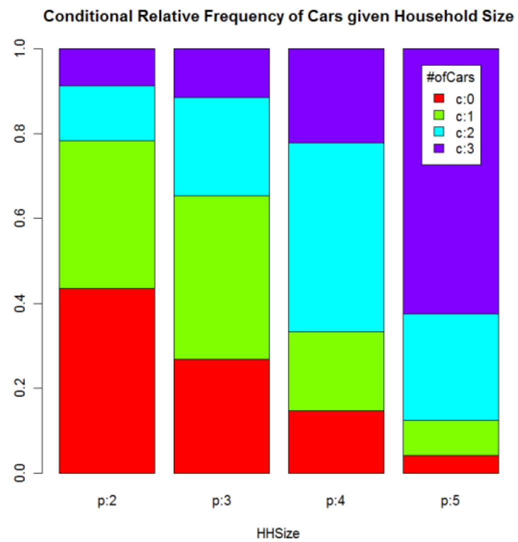
- A bivariate plot similar to that in Figure 5-13 can be generated:



- The column percentages are $f_{i|j} = n_{ij}/n_{+j}$. These can be viewed as conditional probabilities $\Pr(C = i|P = j) = \frac{\Pr(C=i \cap P=j)}{\Pr(P=j)}$ with the marginal probability is

```
> addmargins(prop.table(hh.car.tab,2),1)
     HHSize
cars         p:2        p:3        p:4        p:5
  c:0 0.43478261 0.26923077 0.14814815 0.04166667
  c:1 0.34782609 0.38461538 0.18518519 0.08333333
  c:2 0.13043478 0.23076923 0.44444444 0.25000000
  c:3 0.08695652 0.11538462 0.22222222 0.62500000
  Sum 1.00000000 1.00000000 1.00000000 1.00000000
```

- Plot stacked plot of the column percentages given the household size



- An analog procedure can be applied to calculate row percentages and plot the household size percentage give the number of cars.
  However, this does not make sense in this example because one rarely can assume that the number of cars influences the household size.
- In general terms:
  - For population probabilities the row and column indices become random variables I and J.
    The joint probability of a particular cell i, j simply is $\Pr(I = i \cap J = j)$.
    Relationship to the Multinomial Distribution: In many cases one can assume that the cell

counts n_{ij} in a contingency table follow a multinomial distribution with theoretical population cell probabilities $\pi_{ij} = \Pr(I = i \cap J = j)$ and the total number of experiments being $n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$.

- A ***marginal*** column probability is $\Pr(J = j) = \sum_i \Pr(I = i \cap J = j)$.

  Because the elementary outcomes $(I = i \cap J = j)$ are assumed to be independent we are allowed to sum these. Analogue for the marginal row probabilities

  $\Pr(I = i) = \sum_j \Pr(I = i \cap J = j)$

- The total probability sums to one according to the Kolmogorov's axioms:

  $\sum_j \sum_i \Pr(I = i, J = j) = 1$.

- The conditional probability is $\Pr(I = i \mid J = j) = \dfrac{\Pr(I = i, J = j)}{\Pr(J = j)}$

The Covariance between Pairs of Variables

- The covariance between metric pairs of variables is defined for discrete and continuous random variables.
  - Discrete case: $Cov(I, J) = \sum_i \sum_j \big(i - E(I)\big) \cdot \big(j - E(J)\big) \cdot \Pr(I = i \cap J = j)$

- Continuous case: $Cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\cdot\left(x - E(X)\right)\cdot\left(y - E(Y)\right)\cdot dx\cdot dy$ where $f(x,y)$ is the joint density of the random variables $X = x$ and $Y = y$.

- The covariance equations are closely related to the numerator of the correlation coefficient:
  - However, the sample correlation coefficient assumes a ***equal weight*** of $\Pr(I = i, J = j) \Leftrightarrow 1/n$.
  - Whereas the population covariance has a ***probability weight*** of $\Pr(I = i \cap J = j)$ for a discrete distribution or $f(x,y)$ for a continuous distribution
- A covariance of zero implies that a ***pair random variables*** is linearly stochastically independent, $\Pr(I,J) = \Pr(I)\cdot\Pr(J)$ or $f(X,Y) = f(X)\cdot f(Y)$

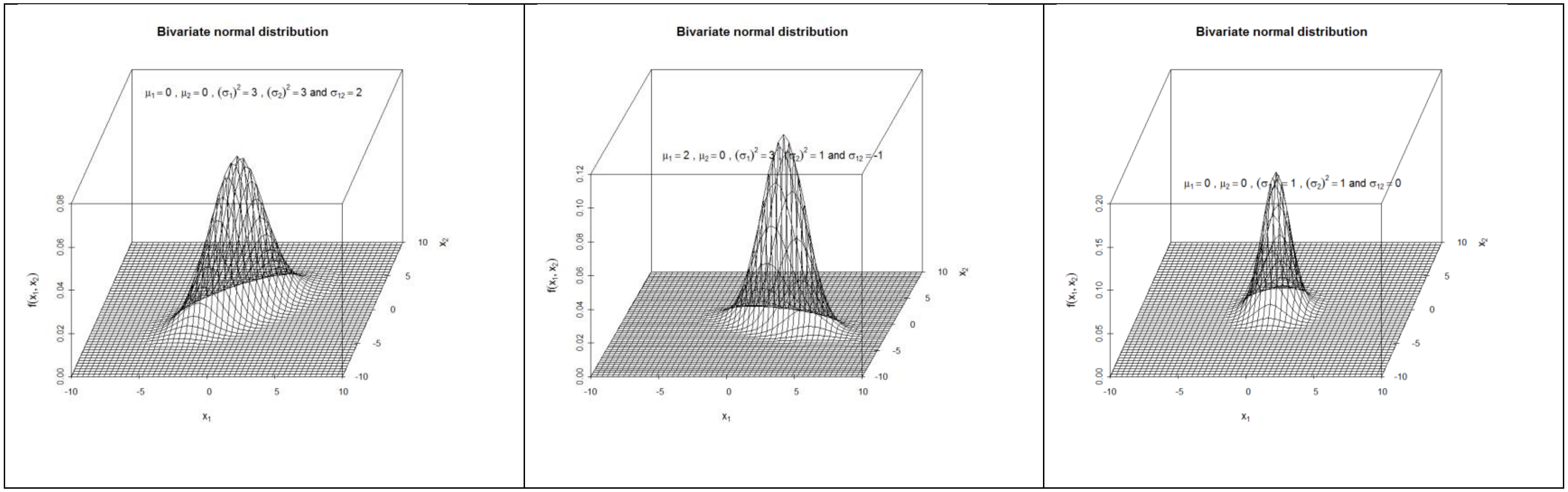
- Pairwise independence, however, *does **not imply*** that three or more random variables are jointly independent.
  Definition of jointly independence for three random variables:
  $\Pr(H,I,J) = \Pr(H)\cdot\Pr(I)\cdot\Pr(J)$ or $f(X,Y,Z) = f(X)\cdot f(Y)\cdot f(Z)$

Continuous Distributions: The bivariate normal distribution (skipped)

- The bivariate normal distribution between two continuous random variables $X_1 \in [-\infty,\infty]$ and $X_2 \in [-\infty,\infty]$ is characterized by five parameters:
  - μ_{X_1} and μ_{X_2} for the central tendency on each axis X_1 and X_2.

- σ_{X_1} and σ_{X_2} for the spread along each axis X_1 and X_2.
- σ_{X_1,X_2} for the covariance between both X_1 and X_2.

- 3-D examples of a bivariate normal distribution are:



- The ***conditional density*** is given by $f(x_1 \mid x_2) = \dfrac{f(x_1, x_2)}{f(x_2)}$ is shown by the x_1 and x_2 gridlines.

- The conditional distribution is again normal distributed with the expectation

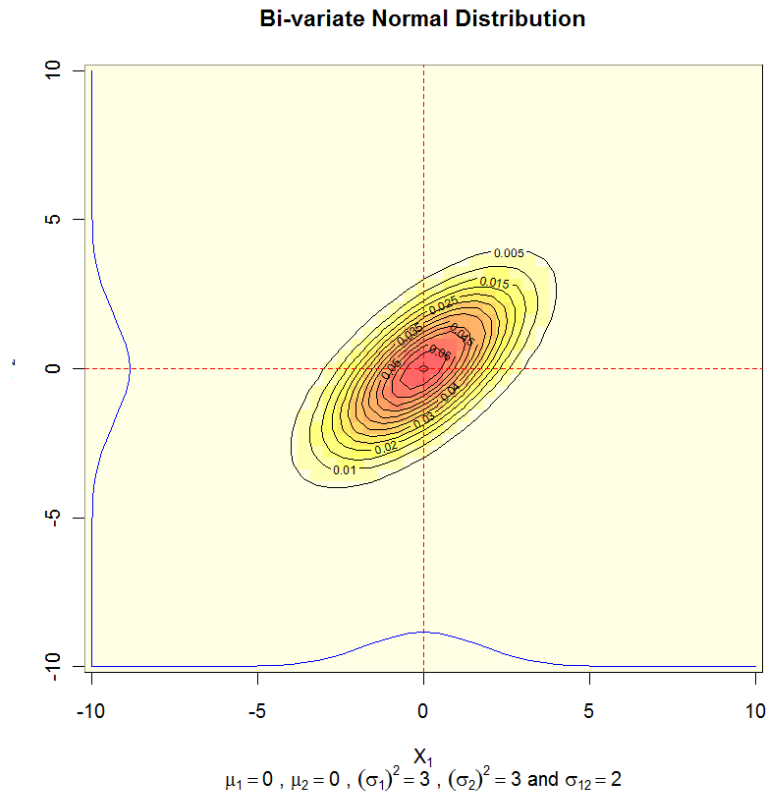
$\mu_{X_1|X_2=x_2} = \mu_{X_1} + \dfrac{\sigma_{X_1,X_2}}{\sigma^2_{X_2}} \cdot \left(x_2 - \mu_{X_2}\right)$ and the variance $\sigma^2_{X_1|X_2=x_2} = \left(1 - \rho^2_{X_1,X_2}\right) \cdot \sigma^2_{X_1}$

- *Contour plot* example with equal density isolines and marginal densities:
- The marginal distributions are again normal distributed with the parameters μ_{X_1} and $\sigma^2_{X_1}$:

  e.g. $N(\mu_{X_1}, \sigma^2_{X_1}) = f(X_1 = x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) \cdot dx_2$

  .

  These marginal distributions are schematically shown by the blue curves at the margins



- The total probability under the density is

$$\int_{-\infty}^{\infty} \underbrace{\left(\int_{-\infty}^{\infty} f(x_1, x_2) \cdot dx_2 \right)}_{= f(x_1)} \cdot dx_1 = 1$$