CS 5135/6035 Learning Probabilistic Models

Lecture 6: Multivariate Probability Distributions [1]

Gowtham Atluri

September 9, 2018

[1] These slides cover material from Chapter 6 of the book, Scheaffer, Richard L., and Linda Young. Introduction to Probability and its Applications, 2009.

Part a. Standard Distributions

- Uniform distribution
- Exponential distribution
- Gamma distribution
- Normal/Gaussian distribution
- Beta distribution
- Weibull distribution

Gamma Distribution

- In case of electronic components
  - few have very short life length
  - many have something close to an average life length
  - very few have extraordinarily long life length

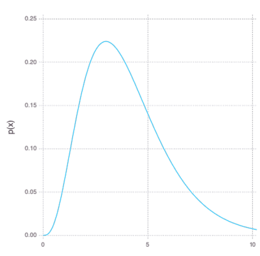
Probability density function

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

where

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

$\Gamma(\alpha)\beta^\alpha$ is a normalizing factor.



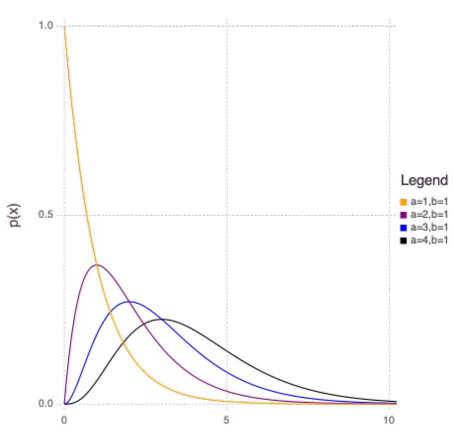
Exponential dist. is a special case of Gamma dist. ($\alpha = 1$).

Parameters: α - shape parameter (Julia code)

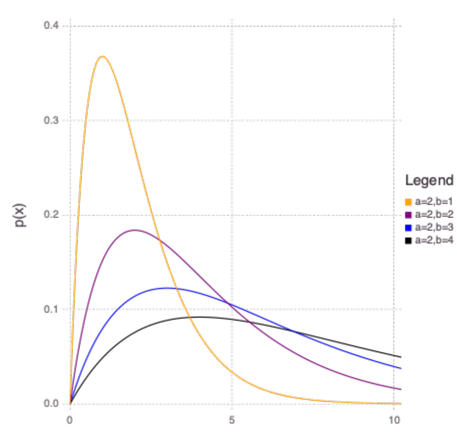
```
xa = collect(0:0.01:20);
gpdfa = pdf.(Gamma(1,1),xa);
gpdfb = pdf.(Gamma(2,1),xa);
gpdfc = pdf.(Gamma(3,1),xa);
gpdfd = pdf.(Gamma(4,1),xa);

myplot = plot(layer(x=xa,y=gpdfa, Geom.line,
                Theme(default_color=colorant"orange")),
            layer(x=xa,y=gpdfb, Geom.line,
                Theme(default_color=colorant"purple")),
            layer(x=xa,y=gpdfc, Geom.line,
                Theme(default_color=colorant"blue")),
            layer(x=xa,y=gpdfd, Geom.line,
                Theme(default_color=colorant"black")),
            Coord.Cartesian(xmin=0, xmax=10),Guide.ylabel("p(x)")
            Guide.manual_color_key("Legend", ["a=1,b=1", "a=2,b=1
                "a=3,b=1","a=4,b=1"], ["orange", "purple","blue",
draw(PNG("./figs/gamma_pdf2.png", 5inch, 5inch), myplot);
```


Parameters: α - shape parameter



Parameters: β - rate parameter



Mean and variance of Gamma distribution

Mean

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$
$$= \int_{0}^{\infty} x \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx = \alpha\beta$$

Variance

$$\sigma^2 = E(x^2) - \mu^2$$
$$= \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$$

Inverse-Gamma Distribution

- If a r.v. $\frac{1}{x}$ follows Gamma distribution with parameters α and β, then x has Inverse-Gamma distribution.
- Generally used in Bayesian analysis

Probability density function

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

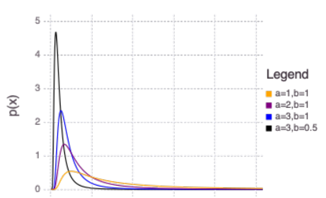
Inverse-Gamma Distribution

- If a r.v. $\frac{1}{x}$ follows Gamma distribution with parameters α and β, then x has Inverse-Gamma distribution.
- Generally used in Bayesian analysis

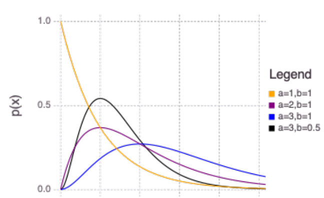
Probability density function

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

$x \sim InverseGamma(\alpha, \beta)$



$x \sim Gamma(\alpha, \beta)$



Chi-square distribution

- Chi-square distribution is a special case of Gamma distribution
  - $\alpha = \frac{k}{2}$ and $\beta = 2$.
- This is also expressed as $\chi^2_k \sim \Gamma(\frac{k}{2}, 2)$

Probability density function

$$f(x) = \begin{cases} \frac{1}{\Gamma(\frac{k}{2})2^{k/2}} x^{\frac{k}{2}-1} e^{-x/2}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

- Chi-square has additive property since the inverse scale parameter is fixed
  - If $x_1 \sim \chi^2_{k_1}$ and $x_2 \sim \chi^2_{k_2}$ are independent χ^2 variables, then $x_1 + x_2 \sim \chi^2_{k_1+k_2}$

Inverse chi-square distribution

- Inverse chi-square distribution is a special case of Inverse Gamma distribution
  - $\alpha = \frac{k}{2}$ and $\beta = 2$.

Probability density function

$$f(x) = \begin{cases} \frac{1}{\Gamma(\frac{k}{2})2^{\frac{k}{2}}} x^{-(\frac{k}{2}-1)} e^{-1/2x}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

Julia functions: Gamma

```
# params \alpa and \beta
d = Gamma(1,2)
```

```
## Distributions.Gamma{Float64}( =1.0,  =2.0)
```

```
# params \alpha and \beta^-1
d = InverseGamma(1,0.5)
```

```
## Distributions.InverseGamma{Float64}(
## invd: Distributions.Gamma{Float64}( =1.0,  =2.0)
## 	: 0.5
## )
```

Julia functions: Chisquared

```
# param k
d = Chisq(1)
```

```
## Distributions.Chisq{Float64}( =1.0)
```

- No InverseChisq function in Julia, but InverseGamma can be used to sample from this distribution.

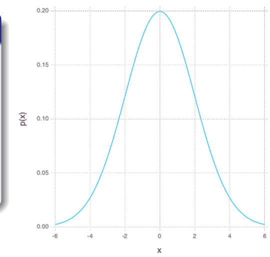
Gaussian/Normal Distribution

- Many naturally occurring measurements (e.g., heights of men) tend to have a relative freq. dist.
  - with some small values
  - with most values close to the average
  - with some high values, resulting in a bell shaped symmetric curve
- The most widely used probability distribution

Probability density function

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

where μ is the mean value and σ^2 is the variance (spread).



Mean and variance of Gaussian distribution

Mean

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$
$$= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \mu$$

Variance

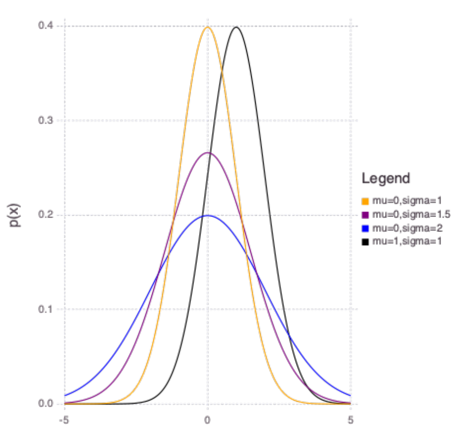
$$E[(x-\mu)^2] = \sigma^2$$

Parameters: μ and σ

```
xa = collect(-5:0.01:5);
gpdfa = pdf.(Normal(0,1),xa);
gpdfb = pdf.(Normal(0,1.5),xa);
gpdfc = pdf.(Normal(0,2),xa);
gpdfd = pdf.(Normal(1,1),xa);

myplot = plot(layer(x=xa,y=gpdfa, Geom.line,
                    Theme(default_color=colorant"orange")),
              layer(x=xa,y=gpdfb, Geom.line,
                    Theme(default_color=colorant"purple")),
              layer(x=xa,y=gpdfc, Geom.line,
                    Theme(default_color=colorant"blue")),
              layer(x=xa,y=gpdfd, Geom.line,
                    Theme(default_color=colorant"black")),
              Coord.Cartesian(xmin=-5, xmax=5),Guide.ylabel("
              Guide.manual_color_key("Legend", ["mu=0,sigma=1'
                 "mu=0,sigma=2","mu=1,sigma=1"], ["orange", "
```


Parameters: μ and σ



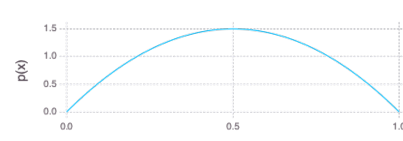
Beta Distribution

- Exponential, Gamma, and Normal distributions are positive over an infinite interval
- Beta distribution is constrained to the interval (0,1)

Probability density function

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

where α and β are positive constants.
$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is a normalizing factor.



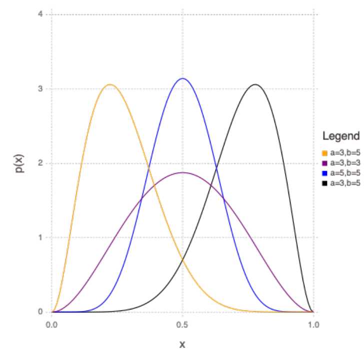
Beta Distribution: Julia code

```
xa = collect(0:0.01:1);
gpdfa = pdf.(Beta(3,8),xa);
gpdfb = pdf.(Beta(3,3),xa);
gpdfc = pdf.(Beta(8,8),xa);
gpdfd = pdf.(Beta(8,3),xa);

myplot = plot(layer(x=xa,y=gpdfa, Geom.line,
                    Theme(default_color=colorant"orange")),
              layer(x=xa,y=gpdfb, Geom.line,
                    Theme(default_color=colorant"purple")),
              layer(x=xa,y=gpdfc, Geom.line,
                    Theme(default_color=colorant"blue")),
              layer(x=xa,y=gpdfd, Geom.line,
                    Theme(default_color=colorant"black")),
              Coord.Cartesian(xmin=0, xmax=1),Guide.ylabel("p(x)"),
              Guide.manual_color_key("Legend", ["a=3,b=5", "a=3,b=3'
                    "a=5,b=5","a=3,b=5"], ["orange", "purple","blue'
draw(PNG("./figs/beta_pdf2.png", 5inch, 5inch), myplot);
```


Beta Distribution

- A rich distribution that can model a range of shapes



Beta Distribution: Mean and Variance

$$
\begin{aligned}
E(x) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_0^1 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

Similar manipulations reveal,

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Weibull Distribution

- Gamma dist. is used to model life lengths of components
  - failure rate function for Gamma dist. has an upper bound
  - this limits applicability to real systems
- Weibull dist. provides a better distribution for life length data

Probability density function

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

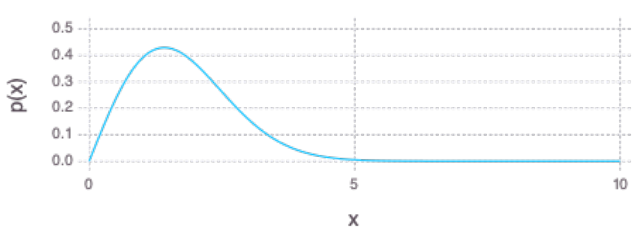
where θ and γ are positive parameters.

- For $\gamma = 1$, this becomes exponential dist.
- For $\gamma > 1$, the function looks like gamma functions, with different mathematical properties

$$cdf(x) = \begin{cases} 0, & x < 0 \\ \int_0^x \frac{\gamma}{\theta} t^{\gamma-1} e^{-t^{\gamma}/\theta} dt = -e^{-t^{\gamma}/\theta}|_0^x = 1 - e^{-x^{\gamma}/\theta}, & x \geq 0 \end{cases}$$

Weibull Distribution: Plot pdf using Julia

```
d = Weibull(2, 2);
xax = collect(0:0.01:10);
gpdf = pdf.(d,xax);
myplot = plot(x=xax,y=gpdf, Geom.line,
              Coord.Cartesian(xmin=0, xmax=10),Guide.ylabel("I
draw(PNG("./figs/weibull_pdf1.png", 5inch, 2inch), myplot);
```



Weibull Distribution: Mean and Variance

$$E(x) = \int_0^{\infty} x \frac{\gamma}{\theta} x^{\gamma-1} e^{-x^{\gamma}/\theta} dx$$

Let $y = x^{\gamma}$ or $x = y^{1/\gamma}$

$$
\begin{aligned}
E(x) = E(y^{1/\gamma}) &= \int_0^{\infty} y^{1/\gamma} \frac{1}{\theta} e^{-y/\theta} dy \\
&= \frac{1}{\theta} \int_0^{\infty} y^{1/\gamma} e^{-y/\theta} dy \\
&= \frac{1}{\theta} \Gamma(1 + \frac{1}{\gamma}) \theta^(1 + 1/\gamma) \\
&= \theta^{1/\gamma} \Gamma(1 + \frac{1}{\gamma})
\end{aligned}
$$

Bivariate and Multivariate Prob. Dist.

- Univariate vs. Multivariate r.v.s
- Joint probability
  - Discrete
  - Continuous
- Cumulative distribution function
- Marginal probability
- Conditional probability
- Independent random variables

Bivariate and Multivariate Prob. Dist.

- We examined experiments that produced single numerical response
  - life length x of a battery
  - strength y of a steel casing
- Often we want to study the joint behavior of two or more random variables
  - joint behavior of life length and casing strength for batteries
  - to identify a region with a combination of life length and casing strength that is optimal in balancing cost of manufacturing with customer satisfaction
  - this is an example of bivariate distribution
- Other examples
  - a physician studies joint behavior of pulse and exercise
  - an educator studies joint behavior of grades and time devoted to study
  - an economist studies joint behavior of business volume and profits

Example: Titanic survivors

Status	Survivors	Fatalities	Total
First class	203	122	325
Second class	118	167	285
Third class	178	528	706
Crew	212	673	885
Total	711	1490	2201

$$x = \begin{cases} 0, & \text{if passenger survived} \\ 1, & \text{if passenger did not survive} \end{cases}$$

$$y = \begin{cases} 1, & \text{if passenger was in } 1^{st} \text{ class} \\ 2, & \text{if passenger was in } 2^{nd} \text{ class} \\ 3, & \text{if passenger was in } 3^{rd} \text{ class} \\ 4, & \text{if passenger was a crew} \end{cases}$$

Joint probability:

p(x,y)	x= 0	x=1	
y=1	0.09	0.6	0.15
y=2	0.05	0.08	0.13
y=3	0.08	0.24	0.32
y=4	0.10	0.30	0.40
	0.32	0.68	1.00

Joint Probability Distribution: Discrete case

- Let X and Y be discrete random variables. The **joint probability distribution** of x and y is given by

$$p(x, y) = P(X = x, Y = y)$$

  defined for all states x and y.
- All joint probability functions must satisfy:
  1. $p(x, y) \geq 0, \quad x, y \in (R)$
  2. $\sum_x \sum_y p(x, y) = 1$
- The **cumulative distribution function** is defined as

$$cdf(x, y) = P(X \leq x, Y \leq y), \qquad (x, y) \in \mathbb{R}^2$$

$$cdf(x, y) = \sum_{x=-\infty}^{a} \sum_{y=-\infty}^{b} p(x, y)$$

Joint Probability Distribution: Continuous case

- Let X and Y be continuous random variables.
- Let $f(x, y)$ be a bivariate function which forms a probability surface in three dimensions.
- The probability that x lies in one interval and that y lies in another interval is represented as a volume under this surface

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y) dxdy$$

- The cumulative distribution function is

$$cdf(a, b,) = P(X \leq a, Y \leq b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y) dxdy$$

Example

Scenario: A certain process for producing an industrial chemical yields a product that contains two types of impurities.

- Let x denote the proportion of total impurities in the sample.
- Let y denote the proportion of type I impurity among all impurities.
- Joint distribution of x and y is given as

$$f(x, y) = \begin{cases} 2(1 - x), & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

- Calculate the probability that $x \leq 0.5$ and $0.4 \leq y \leq 0.7$

Example

- Joint distribution of x and y is given as

$$f(x,y) = \begin{cases} 2(1-x), & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

- Calculate the probability that $x \le 0.5$ and $0.4 \le y \le 0.7$

$$\begin{aligned} p(0 \le x \le 0.5, 0.4 \le y \le 0.7) &= \int_{0.4}^{0.7} \int_{0}^{0.5} 2(1-x)dxdy && (1) \\ &= \int_{0.4}^{0.7} [-(1-x)^2]_0^{0.5} dy && (2) \\ &= \int_{0.4}^{0.7} 0.75 dy && (3) \\ &= 0.75y|_{0.4}^{0.7} && (4) \\ &= (0.75)(0.3) && (5) \\ &= 0.225 && (6) \end{aligned}$$

Marginal Probability Distribution

Marginal probability function of x and y is given by

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

and

$$f(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Example

Scenario: A certain process for producing an industrial chemical yields a product that contains two types of impurities.

- Let x denote the proportion of total impurities in the sample.
- Let y denote the proportion of type I impurity among all impurities.
- Joint distribution of x and y is given as

$$f(x,y) = \begin{cases} 2(1-x), & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

- Derive marginal probability functions for x and y.

Example

- Joint distribution of x and y is given as

$$f(x,y) = \begin{cases} 2(1-x), & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

- Derive marginal probability functions for x and y.

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= \int_{0}^{1} 2(1-x)dy \\ &= 2(1-x)y|_0^1 \\ &= \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases} \end{aligned}$$

$$\begin{aligned} f(y) &= \int_{-\infty}^{\infty} f(x,y)dx \\ &= \int_{0}^{1} 2(1-x)dx \\ &= -(1-x)^2|_0^1 \\ & \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases} \end{aligned}$$

Conditional Probability Distributions: $p(x|y) = \frac{p(x,y)}{p(y)}$

- Discrete case: Conditional distribution of x, given $y = 1$

$$p(x = r|y = 1) = \frac{p(x = r, y = 1)}{p(y = 1)}$$

  for all states r of x.
- Continuous case
  - Let x and y be continuous r.vs. with joint pdf f(x,y) and marginals f(x) and f(y), respectively.
  - Conditional pdf of x given y is defined by

$$f(x|y) = \begin{cases} \frac{f(x,y)}{f(y)}, & \text{for } f(y) > 0 \\ 0, & \text{elsewhere} \end{cases}$$

  - Conditional pdf of y given x is defined by

$$f(y|x) = \begin{cases} \frac{f(x,y)}{f(x)}, & \text{for } f(x) > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Example

Scenario: A certain process for producing an industrial chemical yields a product that contains two types of impurities.

- Let x denote the proportion of total impurities in the sample.
- Let y denote the proportion of type I impurity among all impurities.
- Joint distribution of x and y is given as

$$f(x,y) = \begin{cases} 2(1-x), & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

- Evaluate the probability that the proportion of type I impurity is less than 0.5, given that the total impurities in the sample is 0.2.
- $p(y < 0.5|x = 0.2)$.

Example

Joint distribution of x and y is given as

$$f(x,y) = \begin{cases} 2(1-x), & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

We know marginal probabilities

$$f(x) = \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases} \qquad f(y) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

- Compute $p(y < 0.5 | x = 0.2)$.

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{2(1-x)}{2(1-x)}$$

$$= \begin{cases} 1, & 0 \le x, y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

Example

$$f(y|x) = \begin{cases} 1, & 0 \le x, y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

We know that $x = 0.2$, but f(y|x) does not depend of x.

$$f(y|x = 0.2) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

Probability of interest $p(y < 0.5 | x = 0.2)$

$$p(y < 0.5 | x = 0.2) = \int_{-\infty}^{\infty} f(y|x = 0.2)dx = \int_{0}^{0.5} 1dy = 0.5.$$

Independent random variables: $p(x,y) = p(x)p(y)$

- Discrete case: Two discrete r.vs x and y are independent, if and only if, for all states r and s of variables x and y,

  $$p(x = r, y = s) = p(x = r)p(y = s)$$

  for all states r of x.
- Continuous case
  - Continuous r.vs. x and y are said to the independent if

    $$f(x,y) = f(x)f(y)$$

    for all values of x and y.

Example

Given continuous r.vs x and y with joint density function

$$f(x,y) = \begin{cases} 2(1-x), & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

Determine if they are independent.

We know marginal probabilities

$$f(x) = \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases} \qquad f(y) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

It is easy to see that $f(x,y) = f(x)f(y)$ for all values of x and y.

Therefore, x and y are independent.

Expected values

- If x and y are discrete r.vs, and $g(x,y)$ is any real-valued functions, the expected value of $g(x,y)$ is

  $$E[g(x,y)] = \sum_x \sum_y g(x,y)p(x,y)$$

  The sum is over all values of (x,y) for which $p(x,y) > 0$
- If x and y are continuous r.vs, and $f(x,y)$ is a joint probability density function

  $$E[g(x,y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

Expected values

- If x and y are independent with means μ_x and μ_y, then

  $$E(xy) = E(x)E(y)$$

- If x and y are independent, g is a function of x alone and h is a function f y alone, then

  $$E[g(x)h(y)] = E[g(x)]E[h(y)]$$

Covariance

- Covariance helps to assess the relationship between two variables.
- Two variables have positive covariance
  - If y tends to be large when x tends to be large
  - If y tends to be small when x tends to be small
- Two variables have positive covariance
  - If y tends to be small when x tends to be large
  - If y tends to be large when x tends to be small

Covariance

Covariance between two random variables x and y is given by

$$cov(x, y) = E[(x - \mu_x)(y - \mu_y)]$$

where $\mu_x = E(x)$ and $\mu_y = E(y)$

- This can also be expressed as $cov(x, y) = E(xy) - \mu_x \mu_y$

Correlation

- Covariance depends on the units of measurement
  - Covariance is found to be 0.2meter^2
  - If we report it in centimeters, it will be 200cm^2
- We need a measure that allows us to judge the strength of the association regardless of the units.

Correlation

Correlation between two random variables x and y is given by

$$\rho = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sqrt{\sigma_x^2\ \sigma_y^2}} = \frac{cov(x, y)}{\sqrt{\sigma_x^2\ \sigma_y^2}} = \frac{cov(x, y)}{\sigma_x\ \sigma_y}$$

- It is a unitless quantity that takes on values between -1 and $+1$.
- If x and y are independent r.vs. Then

$$cov(x, y) = E(xy) - E(x)E(y) = E(x)E(y) - E(x)E(y) = 0$$