Lesson 10

Continuous Random Variables

One may imagine that, given that a discrete r.v. can take discrete values, a continuous random variable is a r.v. that can take a continuum of values; however, the definition of a continuous r.v. is subtle.

It is not enough for a r.v. to have a continuous range to be called continuous.

Def: Assume that X is a r.v. on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. X is said to be continuous if there exists a non-negative measurable function $f: \mathbb{R} \to [0, +\infty)$ such that:

$\forall B \in \mathcal{B}(\mathbb{R})$ $\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \int$

$= \int_{\mathbb{R}}$

In particular If $B = (-\infty, x]$,

$\mathbb{P}_X(B) =$

Remark: Recall that $f: \mathbb{R} \to \mathbb{R}$

is measurable if $\forall B \in \mathcal{B}(\mathbb{R})$,

$f^{\leftarrow}(B) \in \mathcal{B}(\mathbb{R})$. All continuous

functions on $\mathbb{R}$ are measurable,

but the reverse is not true.

f is called <u>a</u> probability density function (pdf) for X.

Note that in general, the integral $\int_B f(t)\,dt$ of the measurable function f is an abstract integral; however, in most applications, $\int_B f(t)\,dt$ can be calculated as a Riemann Integral, which is the integral that is studied in calculus

In particular, when f is well-behaved (i.e. when its is continuous or has a countable number of discontinuities), we can interprete $\int_{-\infty}^{a} f(t)\,dt$ as a Riemann Integral.

Lemma) If f_X is the pdf of a discrete r.v., $\int_{-\infty}^{+\infty} f_X(t)\,dt = 1$

Proof

To summarize, $f_X : \mathbb{R} \to \mathbb{R}$ is a pdf iff

a) $f_X(t) \geq 0 \quad \forall t \in \mathbb{R}$

b) f_X is measurable

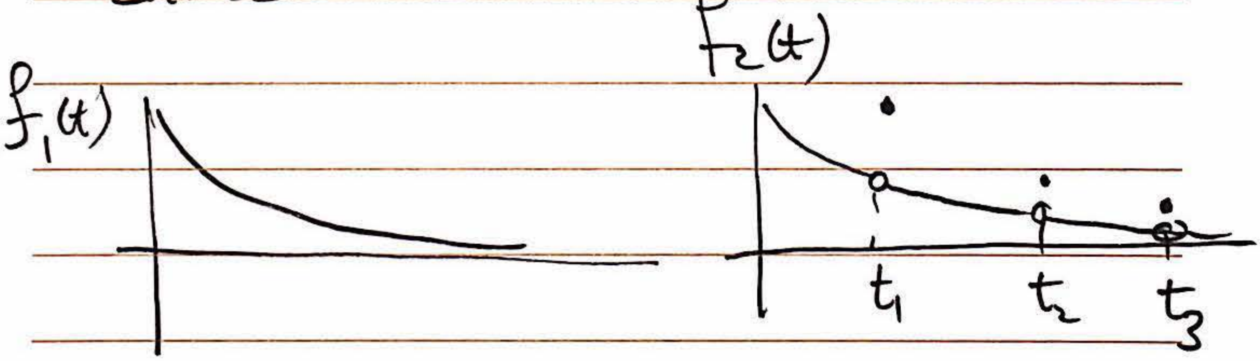
c) $\int_{-\infty}^{+\infty} f_X(t)\, dt = 1$

For any density function, we can find the CDF $F_X = \int_{-\infty}^{x} f_X(t)\, dt$

If the CDF is differentiable at $x = x_0$

$F_X'(x_0) =$

Remark: pdfs are not unique for a random variable. For example, if a pdf is changed at a finite set of points, its integral (cdf) remains the same.



$$\forall\, t \neq t_1, t_2, t_3 \quad f_1(t) = f_2(t)$$

$$\int_{-\infty}^{x} f_1(t) = \int_{-\infty}^{x} f_2(t)\, dx$$

$$\rightarrow F_1(x) = F_2(x)$$

Rarely is this the cause of any mathematical difficulty; therefore, in the sequel, "the pdf" will be used instead of "a pdf" noting that

we actually refer to a "class of pdfs" rather than a single pdf.

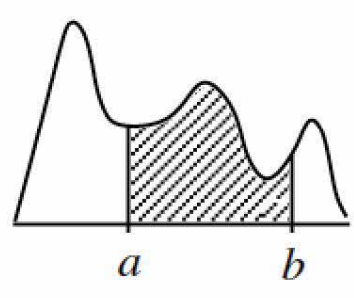
According to the definition of pdf, the probability that X falls in a Borel set B can be calculated as $\int_B f(t)\,dt$.

In particular, if $\forall x$,

$\mathbb{P}(X=x)=0$ (i.e. the CDF is continuous)

$$\mathbb{P}(a<X<b)=\mathbb{P}(a\leq X\leq b)$$

$$=\int$$



Example: A Random variable that is neither continuous nor discrete

X ——— Prob $= \alpha$ $\qquad X$: discrete

$Z \qquad Y$: Continuous

Y ———

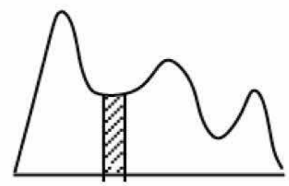
$\mathbb{P}(Z = X) = \alpha \qquad \mathbb{P}(Z = Y) = 1-\alpha$

$F_Z(z) =$

A pdf is similar to a pmf, but probability is encoded as "area" under a pdf, compared to pmfs, for which probabilities are encoded as values.

The pdf of some continuous r.v. X can have values greater than 1, while that's not the case for a pmf. To understand this, observe that

$\mathbb{P}(X \in [x, x+\Delta x]) = \int$



provided that f_X is continuous over $[x, x+\Delta x]$.

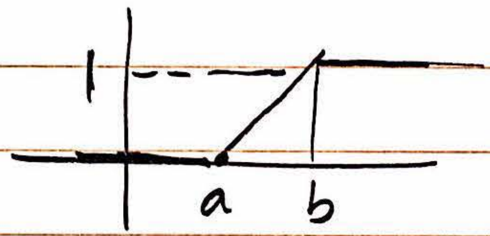
Question: We know that a r.v. is a function $X: \Omega \to \mathbb{R}$ that is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$ measurable. If X is a continuous r.v., does it mean that X is a continuous function?

Important Continuous R.V.'s

a) Uniform

$$X \sim U[a,b]$$

$$F_X(x) = \begin{cases} 0 & x \le a \\ \frac{1}{b-a}(x-a) & a < x \le b \\ 1 & x > b \end{cases}$$



$$f_X(x) = \begin{cases} \frac{d F_X(x)}{dx} = & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

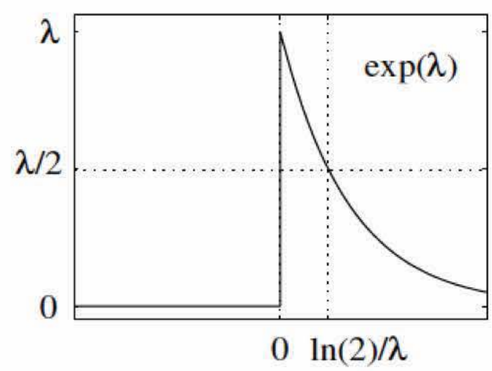
Remark: The values of $f_X(x)$ at $x=a$ and $x=b$ don't matter!

(b) Exponential : $\lambda > 0$

We say $X \sim Exp(\lambda)$ if

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & o.w. \end{cases}$$

Check that it is indeed a CDF.



$$f_X(x) = \begin{cases} & x \geq 0 \\ & x < 0 \end{cases}$$

The Relationship between Geometric and Exponential Random Variables

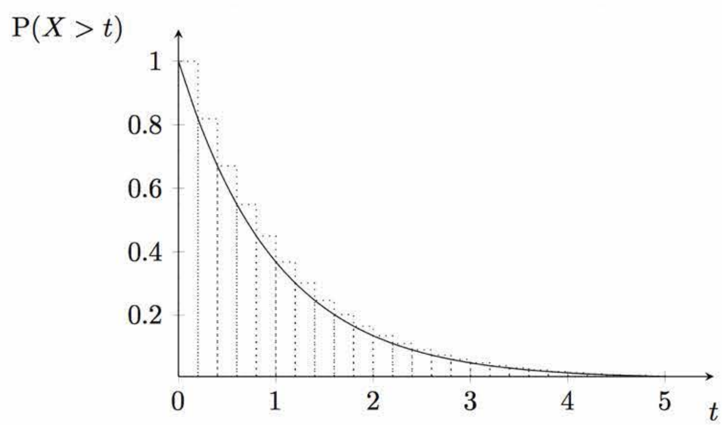
Recall that the number of failures before the first success is observed in a sequence of independent Bernoulli trials is modeled by a $Geo_0(p)$ r.v.:

$$P_X(x) = p(1-p)^x \qquad \forall x \in \{0, 1, 2, \dots\}$$

An exponential random variable can be viewed as a geometric random variable with very small probability of success.

Theorem: Assume that $X_n \sim Geo_0(\frac{\lambda}{n})$, and $X \sim Exp(\lambda)$. Then

$$\lim_{n \to \infty} F_{X_n}(t) = F_X(t)$$



Approximation of the exponential distribution by the geometric distribution.

Note: The exponential r.v. is often used to model "time to a particular event", e.g. "time to failure"

(Recall that a geometric r.v. is used to model "number of trials needed to observe success.)

Memoryless Property

The geometric distribution is memoryless: the number of attempts necessary to see a success is independent of the past attempts.

Lemma:

Assume that $X \sim Geo_1(p)$.

$$\mathbb{P}(X > s+t \mid X > s)$$

$$=$$

Proof:

Exercise: Show that if

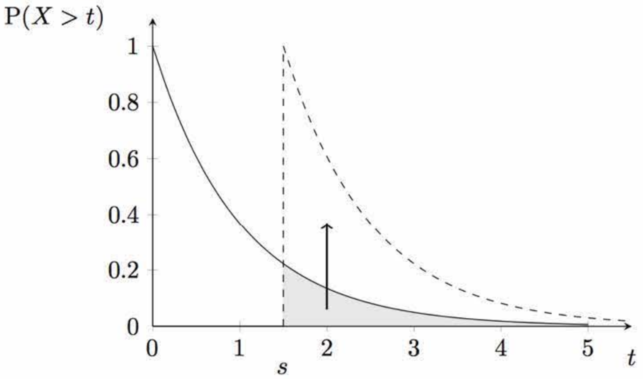
$X \sim Geo_0(p)$, then

$\mathbb{P}(X \geq s+t \mid X \geq s) = \mathbb{P}(X \geq t)$

Memoryless Property for Exponential Random Variables

The Exponential R.V. is a limit of the geometric r.v., and inherits its memoryless property.

Lemma: Assume that $X \sim Exp(\lambda)$, $\lambda > 0$. Then $\mathbb{P}(X > x+t \mid X > x) = \mathbb{P}(X > t)$ $\forall x, t \geq 0$.



Memory-less distribution: the random variable forgets its past.

Proof

Remark: If we interprete $T \sim Exp(\lambda)$ as time to failure of an equipment, the memoryless property says that the event that the equipment continues working for Δt more time is independent from its current age (i.e. the vent that it worked for T_1 periods of time); i.e.

$$\mathbb{P}(T > T_1 + \Delta t \mid T > T_1) = \mathbb{P}(T > \Delta t)$$

Exercise:

Show that if

$f_T(t)$ is nonzero for $t \geq 0$

and T has the memoryless

property, then $T \sim \mathrm{Exp}(\lambda)$.

In other words, being memoryless is

a sufficient condition for being

an exponential r.v., if T is

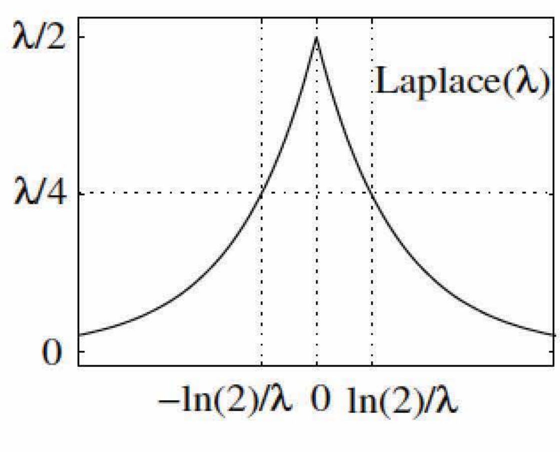
a continuous r.v.

Exercise: Is being memoryless also a sufficient condition for being geometric if X is a discrete r.v.?

(C) The Laplace Distribution

We say X has a Laplace distribution with parameter λ and write $X \sim Laplace(\lambda)$ iff:

$$f_X(x) = \frac{\lambda}{2} e^{-\lambda |x|} \qquad x \in \mathbb{R}$$

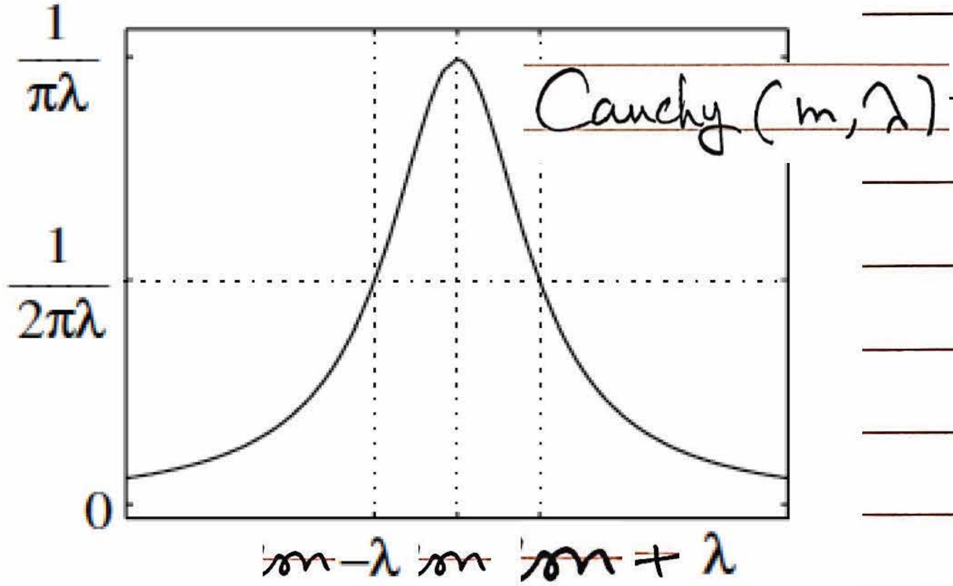


(d) The Cauchy Distribution

We say X has a Cauchy distribution with parameters m and λ and write $X \sim \text{Cauchy}(m, \lambda)$ if

$$f_X(x) = \frac{\lambda/\pi}{\lambda^2 + (x - m)^2} \quad \forall x \in \mathbb{R}$$

Exercise: Show that $f_X(x)$ in above is a pdf.



Remark: $X \sim$ Cauchy$(m=0, \lambda=1)$ is called the standard Cauchy distribution

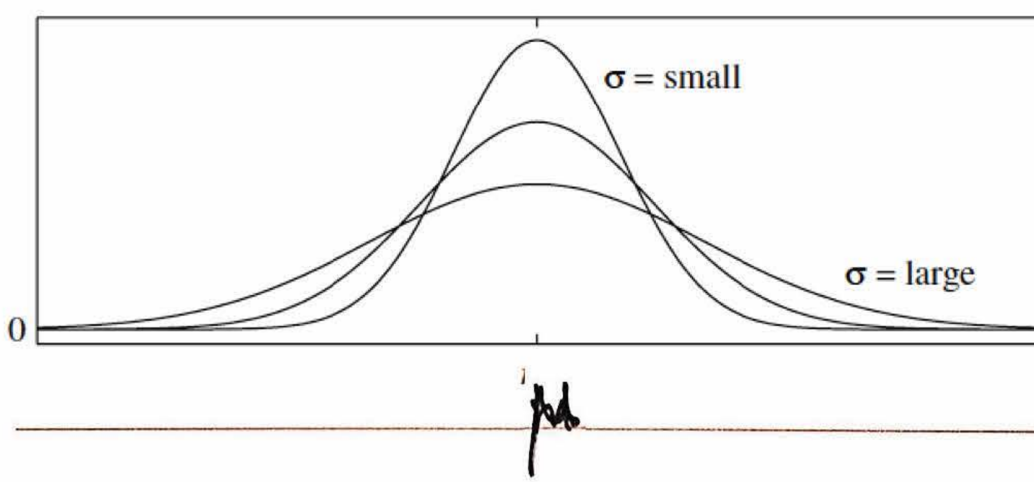
$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2} \quad x \in \mathbb{R}$$

(e) The Normal (Gaussian) Distribution

The Normal Distribution is probably the most widely used and known distribution that appears in many areas of science and engineering, and is of significant interest in Statistics, because of the celebrated Central Limit Theorem.

We say X has a normal (Gaussian) distribution with parameters $\mu \in \mathbb{R}$, $\sigma^2 > 0$ and write $\mathcal{N}(\mu, \sigma^2)$ if

$f_X(x) =$



Note: When $\mu=0$ and $\sigma=1$,

X is called a standard normal

random variable, Z:

$$Z \sim N(0,1)$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Exercise: Show that

$$f_X = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{is a pdf.}$$

Note: when σ becomes larger, the distribution becomes wider. When $\sigma \to 0$, the distribution becomes sharper; therefore $\sigma = 0$ can be seen as a special

limiting case, where the normal r.v. becomes a degenerate r.v. for which $X = \mu$ with probability 1
a.s.

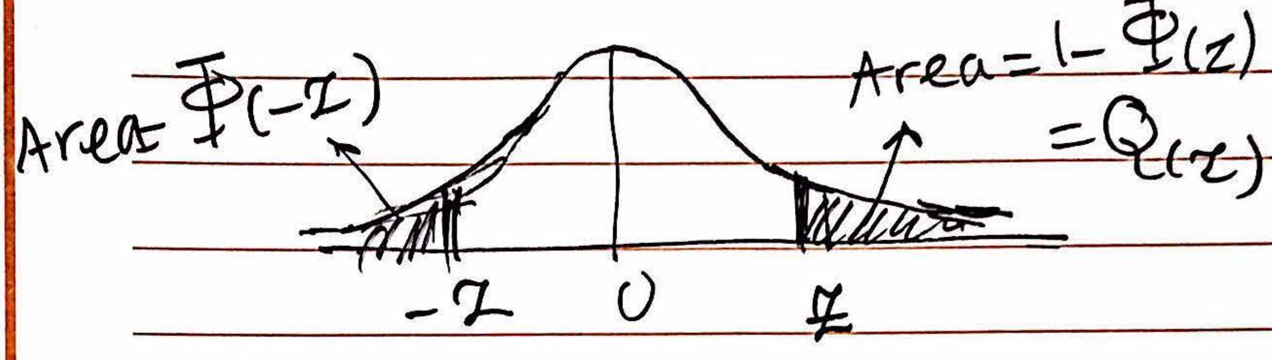
Question: when $\sigma \to 0$, what is the limit of the pdf of $X \sim N(\mu, \sigma^2)$? How do you justify this?

Remark: The CDF of $X \sim \mathcal{N}(\mu, \sigma^2)$ is important in probability and statistics:

$$\Phi_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

Unfortunately, there are no closed-forms for $\Phi_X(x)$; however, there are tables for the CDF of the Standard Normal $\Phi_Z(x)$, $Z \sim \mathcal{N}(0,1)$



By symmetry

$$\Phi(-z) = 1 - \Phi(z) = Q(z)$$

$$Q(z) = \mathbb{P}(Z > z)$$

and $\Phi(z) + Q(z) = 1.$

It can be shown that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then the "standardized" version of X, $Z = \frac{X-\mu}{\sigma}$ is a standard Normal, $Z \sim \mathcal{N}(0,1)$.

Therefore, to calculate $\Phi_X(c)$, one can use the standard Normal table:

$\Phi_X(c) = \mathbb{P}(X \leq c) =$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09

Cumulative Distribution Function, $F(z)$, of the Standard Normal Distribution Table

Exercise: Using the normal table, calculate

$\mathbb{P}(-1.96 \leq Z \leq 1.96)$ and

$\mathbb{P}(-2.58 \leq Z \leq 2.58)$.

(f) The Gamma Distribution and Its Variations

X is said to have a Gamma distribution with parameters λ, p ($X \sim \text{Gamma}(\lambda, p)$

or $X \sim \Gamma(\lambda, p)$) if

$$f_X(x) = \frac{1}{\Gamma(p)} \lambda^p x^{p-1} e^{-\lambda x} \quad \lambda, p \in \mathbb{R}^+, x > 0,$$

where $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx, \; p > 0$

Note: $p \in \mathbb{N} \Rightarrow \Gamma(p) = (p-1)!$

Proof: Integration by parts.

Note: Some authors use a different order for the ~~of~~ parameters.

Special Cases of The Gamma Distribution

$p = 1 \Rightarrow X \sim$

$\lambda = \frac{1}{2}, p = n/2$ for $n \in \mathbb{N}$,

$X \sim \chi_n^2$

χ_n^2 is the chi-squared distribution with n degrees of freedom.

It can be shown that if $X_1, X_2, \ldots, X_n$ are iid and $X_i \sim N(0,1)$, then

if $Y = \sum_{i=1}^{n} X_i^2$, $Y \sim \chi_n^2$.

When $p \in \mathbb{N}$, X is said to be distributed according to an Erlang distribution $X \sim Erlang(p,\lambda)$

It can be shown that if $X_1, X_2, \ldots, X_p$ are iid and $X_i \sim Exp(\lambda)$, $Y = X_1 + X_2 + \cdots + X_p \sim Erlang(p,\lambda)$

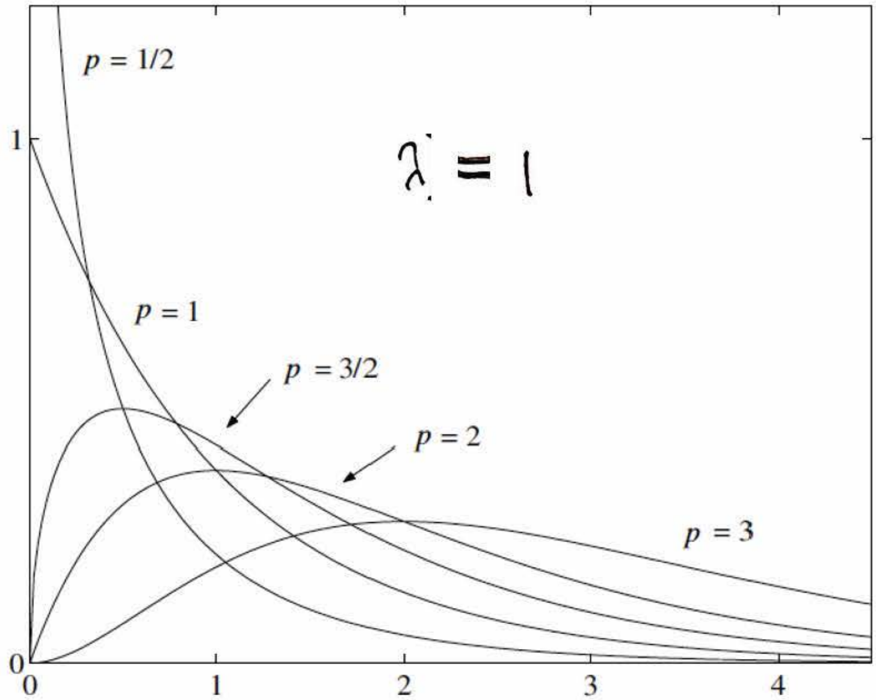


Figure **4.7.** The gamma densities $g_p(x)$ for $p = 1/2$, $p = 1$, $p = 3/2$, $p = 2$, and $p = 3$.

(g) Beta Distribution

X. is said to have a Beta distribution ($X \sim \text{Beta}(a,b)$ $a,b>0$) if:

$$f_X(x) = \frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1}$$

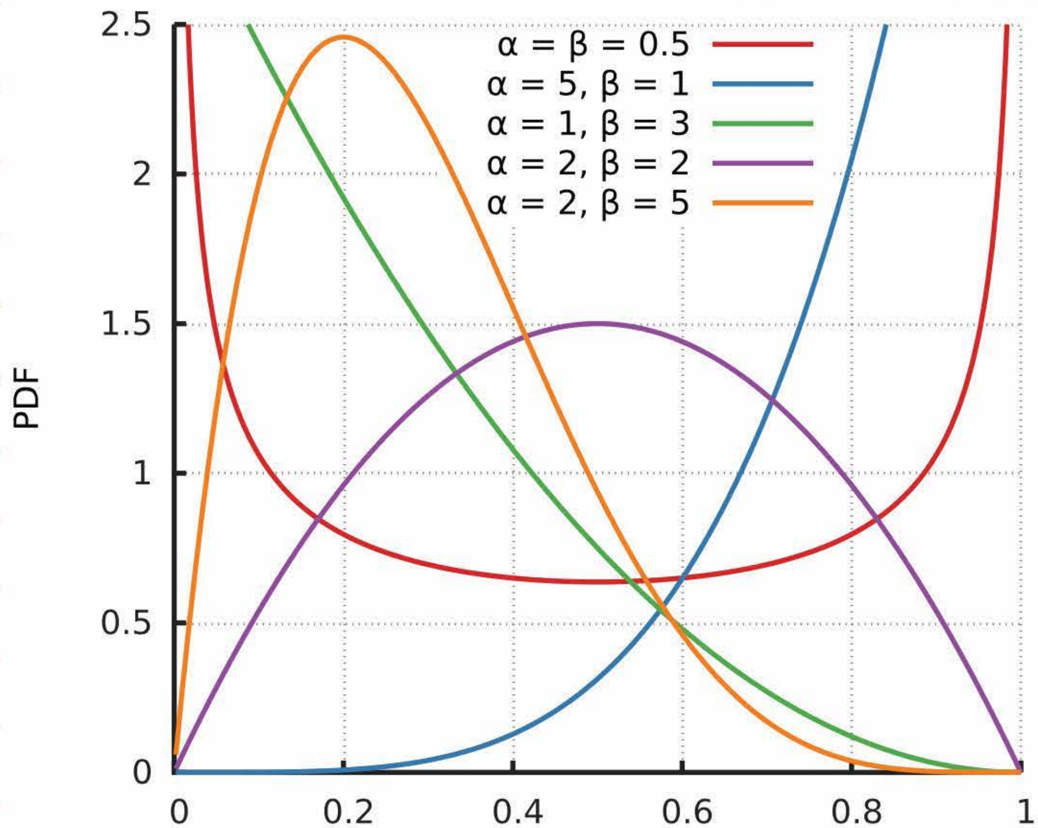
$0 \leq x \leq 1$

Where

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

Proposition: $\beta(a,b) = \dfrac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$

If $a=b=1$, $X \sim$



(h) The Weibull Distribution

The Weibull distribution frequently arises in reliability theory and survival analysis.

X is said to have a Weibull distribution with parameters α, β ($X \sim \text{Weibull}(\alpha, \beta)$) if

$$f_X(x) = \alpha \beta x^{\beta - 1} \exp(-\alpha x^{\beta}) \qquad x \geq 0$$

$$\alpha, \beta > 0$$

Another parameterization of

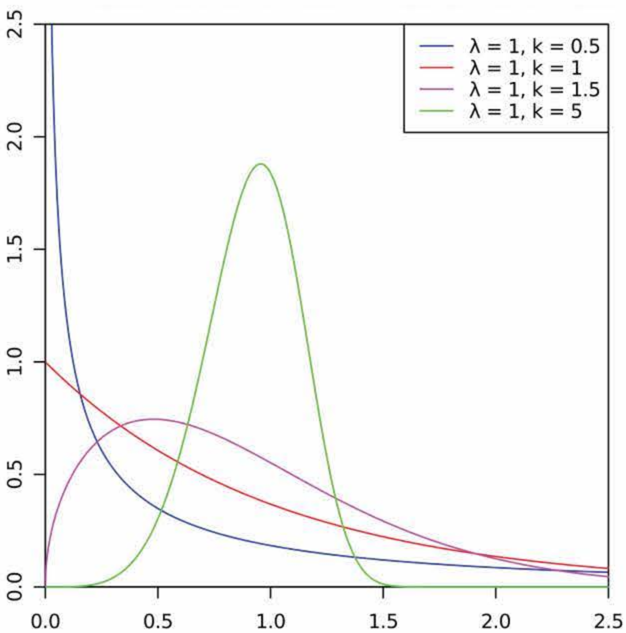
Weibull Distribution

$X \sim \text{Weibull}(\lambda, k) \qquad \lambda, k > 0$

if

$$f_X(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$$

Note that if $\beta = 1$



$X \sim$

(i) Power Laws

Continuous power law distributions are distributions whose "tails" decay proportional to x^{α}, i.e.

$$\mathbb{P}(X > x) = \beta / x^{\alpha} \quad \forall x \geq x_m$$

Therefore, the CDF is

$$F_X(x) = \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X > x)$$

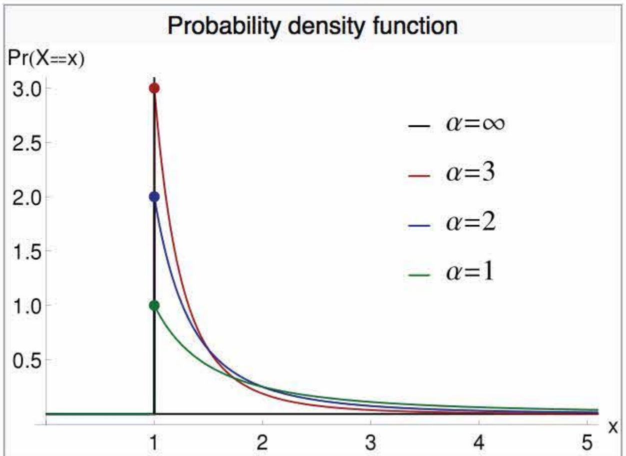
$$= \begin{cases} \\ \end{cases}$$

X has to be continuous, so

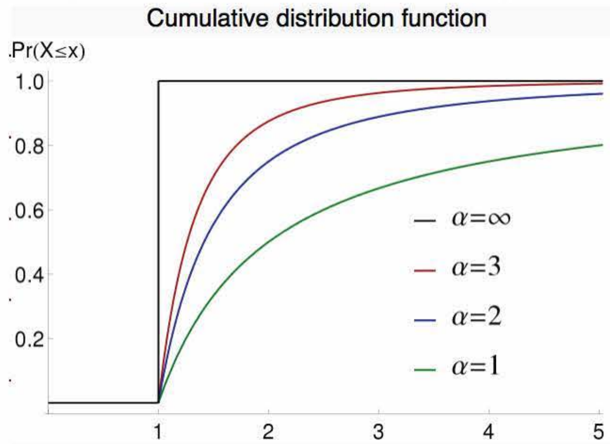
$F_X(x_m) =$

Also, $f_X(x) = \dfrac{dF_X(x)}{dx} =$

Pareto Type I



Pareto Type I probability density functions for various α with $x_m = 1$. As $\alpha \to \infty$ the distribution approaches $\delta(x - x_m)$ where δ is the Dirac delta function.



Pareto Type I cumulative distribution functions for various α with $x_m = 1$.

Location and Scale

Assume that the pdf of a continuous r.v. X is $f_X(x)$.

Then $\lambda f(\lambda(x-c))$ is also a pdf (why?).

λ is called a scale parameter and c is called a location parameter.

Exercise: If $X \sim \text{Laplace}(\lambda)$, plot and interprete $\alpha f_X(\alpha(x-c))$.

Multiple Random Variables and Joint Distributions

Def) Assume that X and Y are r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X, Y are <u>jointly continuous</u>

if there exists a measurable function $f_{X,Y} : \mathbb{R}^2 \longrightarrow [0, +\infty)$

such that

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$$

$=$

We call $f_{X,Y}(x,y)$, the joint pdf of X,Y.

If the cdf is differentiable at (x_0, y_0):

$$\frac{\partial^2 F_{X,Y}(x,y)}{\partial x \, \partial y}\Bigg|_{(x,y)=(x_0,y_0)} = f_{X,Y}(x_0,y_0)$$

If $B \in \mathcal{B}(\mathbb{R}^2)$ is a Borel Set

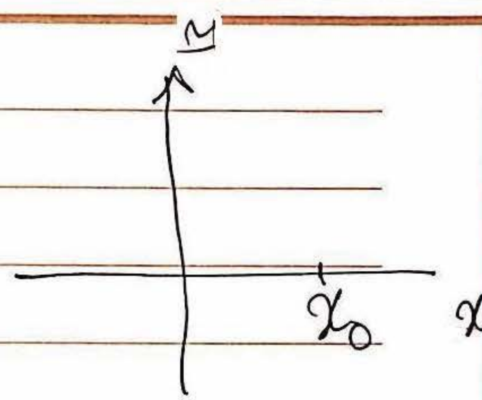
$\mathbb{P}((X,Y) \in B) =$

$=$

Marginal Distributions

Def: If X, Y are jointly continuous r.v.'s, f_X and f_Y are called marginal p.d.f.s of X, Y.

$F_X(x_0) = \mathbb{P}(X \le x_0)$



$=$

Therefore, to find the marginal density, integrate out the unwanted variable.

Important Remark

If X, Y are jointly continuous ~~...~~ we can find pdfs for X and Y; therefore, X is continuous and Y is continuous

The converse is not true.

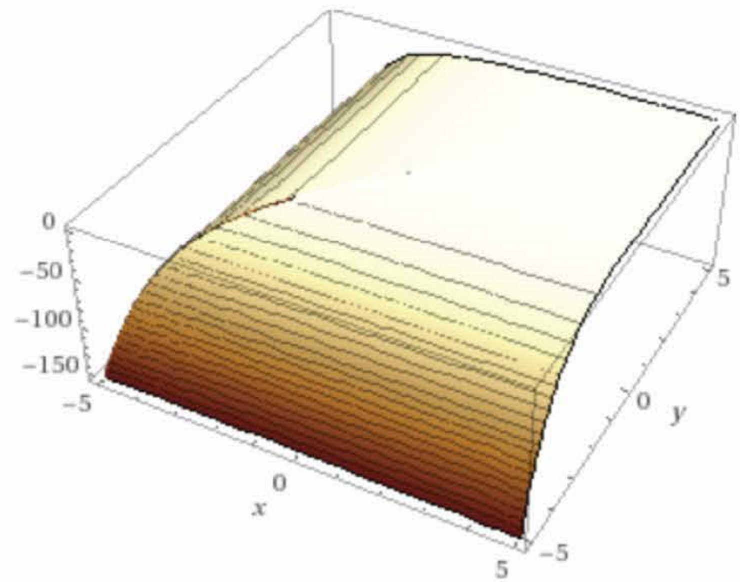
Consider the following example

Example: Let X be a continuous r.v. and let $g: \mathbb{R} \to \mathbb{R}$ be a one-to-one measurable function. Assume that $Y = g(X)$ with probability one Then

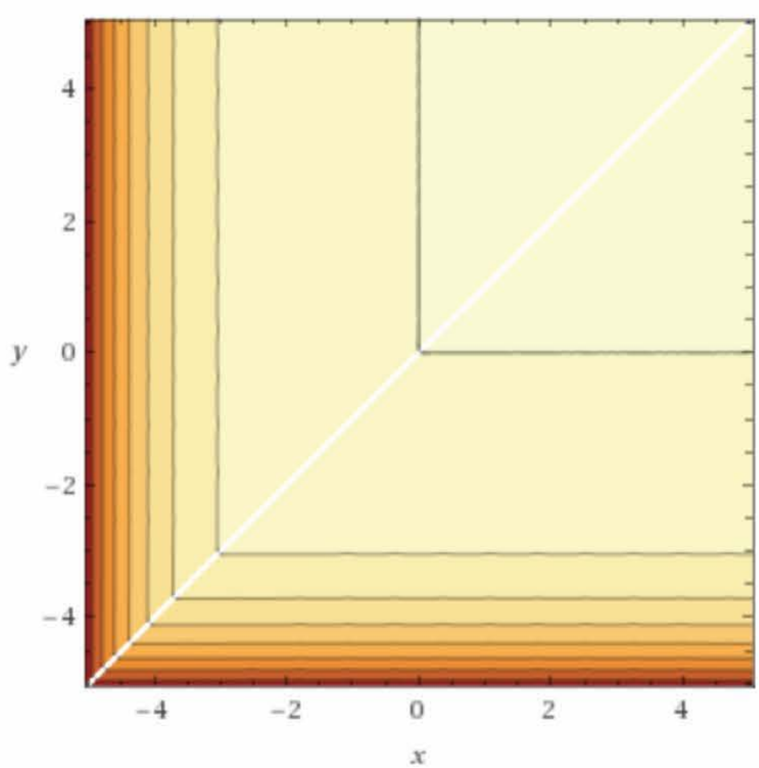
$$F_{X,Y}(x,y) = \left\{ \begin{array}{l} \\ \\ \end{array} \right.$$

plot	$1 - \exp(-\min(x, y))$

3D plot:



Contour plot:



It is called jointly "Singular"

Exercise (Joint CDF):

Assume that $F_{X,Y}(x,y)$ is the joint cdf of X and Y.

Calculate $\mathbb{P}(a < X \leq b, c < Y \leq d)$

Exercise: The joint cdf of X and Y is given:

$$F_{XY}(x,y) = \begin{cases} \dfrac{y + e^{-x(y+1)}}{y+1} - e^{-x} & x,y>0 \\ 0 & \text{otherwise} \end{cases}$$

a) Check that it's indeed a cdf

(b) Find $F_X(x)$, $F_Y(y)$, $f_X(x)$, $f_Y(y)$.

Independence of Continuous Random Variables

Recall X, Y are independent r.v.'s if for all $B_1 \in \mathcal{B}(\mathbb{R})$ and $B_2 \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(X \in B_1, Y \in B_2) = \mathbb{P}(X \in B_1)\mathbb{P}(Y \in B_2)$$

Theorem: Let X, Y be jointly continuous r.v.'s defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The following are equivalent:

(a) X, Y are independent

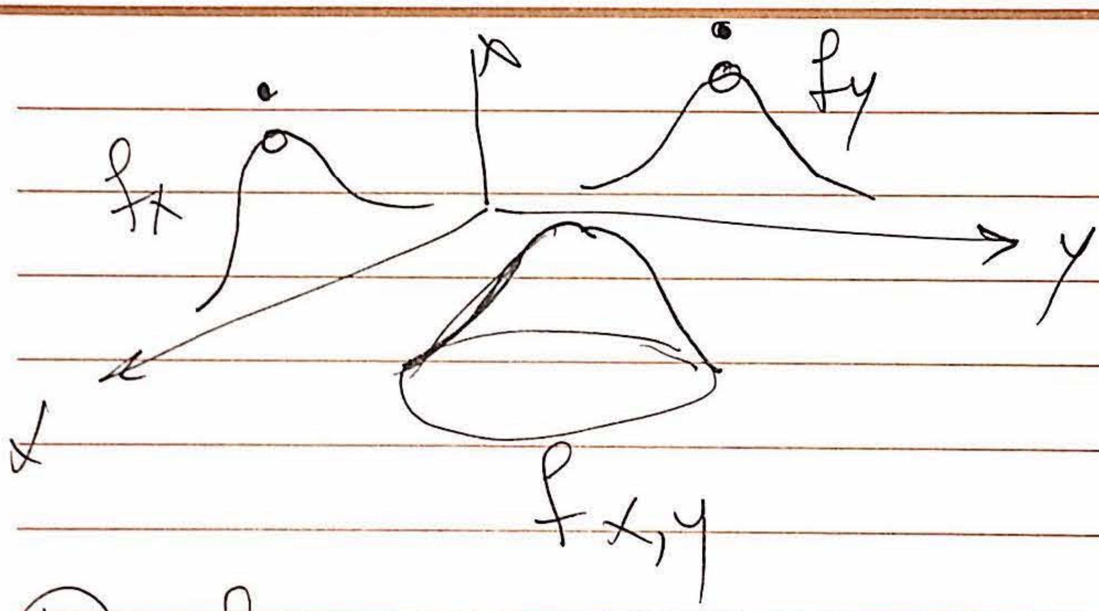
(b) $\forall x, y \in \mathbb{R}$, the events $\{X \leq x\}$, $\{Y \leq y\}$ are independent.

(c) $\forall x, y \quad F_{X,Y}(x,y) = F_X(x) F_Y(y)$

(d) If f_X, f_Y, $f_{X,Y}$ are the joint distributions and marginal distributions, the following equation holds almost everywhere:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

Remark: This means that the inequality may not hold on a set that has Lebesgue measure zero.



Proof: Left to the student

Conditional Probability and Conditional pdfs

Recall that for discrete r.v.'s the conditional pmf of Y given X can be defined as

$$p_{Y|X}(y|x) = \mathbb{P}(Y = y \mid X = x)$$

$$\forall x \text{ s.t } \mathbb{P}(X = x) \neq 0$$

Such a definition will not work for continuous X's, because

$\mathbb{P}(X = x) = 0 \quad \forall x$

Remember that for continuous r.v.'s, probability is encoded as area under the pdf, i.e. $\mathbb{P}(x \leq X \leq x + \Delta x) \approx f_X(x)\,\Delta x$

This inspires us to calculate the conditional cdf of Y given X as

$$F_{Y|X}(y|x) = \lim_{\Delta x \to 0} \mathbb{P}(Y \leq y | \quad)$$

Therefore $f_{Y|X}(y|x) =$

$\forall$ ~~y~~ x s.t. $\boxed{}$

and $f_{Y|X}$ is called the conditional pdf of Y given X

Remark: The previous calculation of $F_{Y|X}$ is valid even if Y is not continuous

The conditional probability of the event $\{Y \in B\}$ given that $X = x$ can therefore be calculated as

$$\mathbb{P}(Y \in B \mid X = x)$$

$$=$$

Exercise: Check that $F_{Y|X}$ is indeed a CDF and has all of the properties of a cdf, i.e.

(a) $\lim_{y \to \infty} F_{Y|X}(y|x) =$

(b) $\lim_{y \to -\infty} F_{Y|X}(y|x) =$

(c) $F_{Y|X}$ is non-decreasing

(d) $F_{Y|X}$ is right-continuous

Moments of Continuous Random Variables

Def (Expectation): Similar to discrete r.v.'s, we define the expectation of a continuous

r.v. X with pdf f_X, as

$\mathbb{E}[X] =$

Also, Similar to the discrete case, if $g: \mathbb{R} \to \mathbb{R}$ is an integrable and measurable function (i.e. if $g(X)$ is a r.v.), the expectation of $Y = g(X)$ can be calculated as:

$$E[Y] = E[g(X)] =$$

This is the law of The Unconscious Statistician (LOTUS) for continuous r.v.'s. Its proof is quite technical and is dropped

Definition: The m^{th} moment and the m^{th} central moment of a continuous r.v. X with pdf f_X are respectively calculated as:

$$\mathbb{E}[X^m] =$$

$$\mathbb{E}[\underbrace{(X - \underbrace{\mathbb{E}[X]}_{\mu})}_{\tilde{X}}{}^m]$$

$$=$$

in particular, when $n=2$

$$\text{var}(X) = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right]$$

$=$

The properties that we learned of expectation are the same as the discrete case. The Definitions of covariance and correlation are also like the discrete case.

Also, our discussions about , the linear algebra aspect of expectations as well as the Cauchy-Schwartz-Bunyakovsky lemma are valid for the

continuous case.

Properties of expectation such as linearity are exactly the same as the continuous case.

Obviously, some results involving Σ's have to be replaced with $\int$:

Lemma: Let X be a non-negative continuous r.v. (i.e. $\mathbb{P}(X<0)=0$ or $X \geq 0$ a.s.). Define $S_X(x) = \mathbb{P}(X > x)$. Then

$$\mathbb{E}[X] = \int_0^{\infty} S_X(x)\,dx$$

Remark: Recall that for a discrete r.v. with non-negative integer values:

$$\mathbb{E}[X] = \sum_{x} \mathbb{P}(X > x)$$

Remark: We have encountered

$$S_X(x) = \mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = 1 - F_X(x)$$

previously when dealing with

exponential r.v.'s. S_X is called the "survival function" of X, because it models "survival" of X. For continuous r.v.'s

$$S_X(x) = \int_x^\infty f_X(t)\,dt$$

Proof

Lemma)

a) The integral

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

is well-defined and finite

if $\int_{-\infty}^{+\infty} |x|\, f_X(x)\, dx < \infty$

in which case we call

X integrable.

b) If only one of $\int_{-\infty}^{0} x\, f_X(x)\, dx$

and $\int_{0}^{\infty} x\, f_X(x)\, dx$ (and not

both) is infinite, $\mathbb{E}[X]$

is infinite, but still well defined.

c) If both $\int_{-\infty}^{0} x f_X(x)\,dx$ and $\int_{0}^{\infty} x f_X(x)\,dx$ are infinite, $\mathbb{E}[X]$ is undefined

Proof: Left as an exercise

Exercise: If $k, m \in \mathbb{Z}^+$ and $k \le m$, then show that if the m^{th} moment of a continuous r.v. is finite, its k^{th} moment is also finite.

Remark: This is also correct for discrete r.v.'s.

Moments of Important

R.v.'s.

a) $X \sim Exp(\lambda)$

$$\mathbb{E}[X^k] = \int_0^\infty x^k \lambda e^{-\lambda x} dx$$

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

$\mu_X = \mathbb{E}[X] =$

$Var(X) =$

(b) $X \sim \text{Gamma}(\lambda, p)$

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

$\mathbb{E}[X^k]$

$\mu_X =$

$\text{Var}(X) = \sigma_X^2 =$

(c) $X \sim \text{Beta}(a, b)$

Recall $\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$

$\mathbb{E}[X^k]$

$$= \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$$

$=$

μ_X :

Var(x)

(d) Uniform $X \sim U(a,b)$

$U(0,1) = \text{Beta}(1,1)$

(d) $X \sim \text{Normal}(\mu, \sigma^2)$

$\mathbb{E}\left[\frac{X-\mu}{\sigma}\right] =$

$\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^k\right]$

$= \mathbb{E}[Z^k] =$

(e) Power Laws (Type I Pareto Distribution)

$$f(x) = \alpha \frac{x_m^{\alpha}}{x^{\alpha+1}} \qquad x \geq x_m$$

$$\mathbb{E}[X^k] =$$

(f) Cauchy

$$X \sim \text{Cauchy}(m, \lambda)$$

$$f_X(x) = \frac{\lambda/\pi}{\lambda^2 + (x-m)^2}$$

$$\mathbb{E}\left[\frac{X-m}{\lambda}\right] =$$

Question: What can be said about higher order moments of Cauchy (m, λ)?

Exercise: Show that when

$X \sim \text{Weibull}(\alpha, \beta)$, i.e.

$$f_X(x) = \alpha\beta\, x^{\beta-1} \exp(-\alpha x^{\beta}),$$

$$\mathbb{E}[X^k] = \frac{k}{\beta\,\alpha^{k/\beta}}\,\Gamma(k/\beta).$$

More on Multiple R.V.'s

The Law of The Unconscious Statistician

Assume that X, Y are jointly continuous random variables with joint pdf $f_{X,Y}$ and $g: \mathbb{R}^2 \to \mathbb{R}$ is measurable, i.e. $Z = g(X,Y)$ is a r.v. Then

$$E[Z] = E[g(X,Y)]$$

$$= \iint$$

Proof: Like the single variable case, it is omitted.

Example: Assume that $Z = X + Y$ where X and Y are independent, with $X \sim N(0,1)$ and $Y \sim Laplace(1)$. Find $Cov(X, Z)$.

The Law of Total Probability

Recall, if $B \in \mathcal{B}(\mathbb{R}^2)$:

$$\mathbb{P}\left((X,Y) \in B \mid X = x\right)$$

$$= \int_B f_{Y|X}(y|x)\,dy$$

$$= \int_{\mathbb{R}} I_B(x,y)\, f_{Y|X}(y|x)\,dy$$

Because $\mathbb{P}(X=x) = 0$, we again need to use the event $\{x \leq X \leq x + \Delta x\}$ to show the following law of Total probability:

$$\mathbb{P}((X,Y)\in B) = \int_{\mathbb{R}} \mathbb{P}((X,Y)\in B \mid X=x) f_X(x)\,dx$$

Proof: Exercise

Exercise: Multiplicative Noise

A random Signal S is communicated over a channel with multiplicative noise N. The received signal is $R = \underset{\sim}{N} S$. Assume that $\underset{\sim}{N}, S$ are jointly continuous with joint pdf $f_{N,S}$. Find the cdf and pdf of R, F_R, f_R.

Question: If X, Y are jointly continuous, how can we formulate the Baye's Rule in terms of their pdfs?

The Substitution Law

Using the same line of reasoning (i.e. considering $\{x \le X \le x+\Delta x\}$ and $\Delta x \to 0$), we have:

$$\mathbb{P}\{(X,Y)\in B \mid X=x\} = \mathbb{P}\{(x,Y)\in B \mid X=x\}$$

Example: Assume that

$Y \sim Exp(\lambda)$ and $X \sim U(a,b)$

and X, Y are independent.

Use Total Probability to

calculate $\mathbb{P}(Y > X)$.

Conditional Expectation

Since $\mathbb{P}(Y \in B \mid X=x)$ is

$$\int_B f_{Y|X}(y|x)\,dy \quad \forall B \in \mathcal{B}(\mathbb{R})$$

we define $\mathbb{E}[Y|X=x]$ as

$$\mathbb{E}[Y|X=x] = \int$$

We can show that The Law of The Unconscious Statistician holds for Conditional Expectation:

$$\mathbb{E}[g(Y)|X=x] =$$

and

$\mathbb{E}[g(X,Y)|X=x]=$

Also, the substitution law holds:

$\mathbb{E}[g(X,Y)|X=x]=\mathbb{E}$

Furthermore, the law of Total expectation is valid in the continuous cases:

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{+\infty} \mathbb{E}[g(X,Y) \mid X=x] f_X(x)\,dx$$

We know that $\mathbb{E}[g(X,Y) \mid X=x]$ is a function of x. Let

$$e(x) = \mathbb{E}[g(X,Y) \mid X=x]$$

Then $e(X)$ is a r.v.

So,

$$\mathbb{E}[e(X)] = \mathbb{E}[g(X,Y)]$$

in case $g(x,y) = y$

$$\mathbb{E}\big[\mathbb{E}[Y|X]\big] = \mathbb{E}[Y]$$

or

$$\mathbb{E}[Y] = \int_{-\infty}^{+\infty} \mathbb{E}[Y|X=x]\, f_X(x)\, dx$$

Exercise: Suppose $X \sim Exp(1)$ and given $X=x$, Y is conditionally normal with mean 0 and variance x^2.

Evaluate $\mathbb{E}[Y^2]$ and $\mathbb{E}[Y^2X^3]$

Remark: It is worth mentioning that the Law of Total Expectation is a general form of the Law of Total Probability. In particular

$$\mathbb{E}[g(x,Y)] = \int_{-\infty}^{+\infty} \mathbb{E}[g(x,Y)|X=x]\, f_X(x)\, dx$$

Using $g(y) = I_B(y)$, we can write:

More Than Two random Variables

If $X_1, \ldots, X_n$ are jointly continuous r.v.'s, their joint cdf can be calculated from a joint pdf:

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \int_{-\infty}^{x} f_{X_1,\ldots,X_n}(x_1',\ldots,x_n')\,d\underline{x}'$$

It's intuitively clear that the joint pdf of any number of them can be calculated by integrating the unwanted variable, e.g. X, Y, Z:

$f_{XZ}(x, z) = \int_{-\infty}^{+\infty}$

and

$f_X(x) =$

Exercise: $f_{XYZ}(x, y, z) = \frac{3z^2}{7\sqrt{2\pi}} e^{-zy} \exp\left[-\frac{1}{2}\left(\frac{x-y}{z}\right)^2\right]$

for $y \geq 0$

and $1 \leq z \leq 2$

Find $f_{YZ}(y, z)$ and $f_{X|Y,Z}(x|y, z)$.

Also find $f_Z(z)$, $f_{Y|Z}(y|z)$, and $f_{X,Y|Z}(x, y|z)$

The Substitution Law and The Law of Total Probability/ Expectation for Three R.V.'s

Assume that X, Y, Z are r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ and are jointly continuous. Then

$$\mathbb{E}[g(X,Y,Z)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y,z) f_{X,Y,Z}(x,y,z)\, dx\, dy\, dz$$

$$\mathbb{E}[g(X,Y,Z) \mid Y=y, Z=z]$$

$$= \int_{-\infty}^{+\infty} g(x,y,z) f_{X|Y,Z}(x|y,z)\, dx$$

Then

$$\mathbb{E}[g(X,Y,Z)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathbb{E}[g(X,Y,Z) \mid Y=y, Z=z]\, f_{Y,Z}(y,z)\,dy\,dz$$

and

$$\mathbb{E}[g(X,Y,Z) \mid Y=y, Z=z] = \mathbb{E}[g(X,y,z) \mid Y=y, Z=z]$$

Exercise: Assume that X, Y, Z are i.i.d r.v.'s distributed according to $U(0,1)$. Find $\mathbb{P}(X \geq YZ)$

Conditional Expectation of Jointly Continuous R.V.'s

Remark: When X, Y are jointly continuous, the random variable $\mathbb{E}[Y|X]$ may be

discrete and is not necessarily continuous. As an example observe that if X, Y are independent and g, h are measurable functions, such that $g(X)$ and $h(Y)$ are jointly continuous,

$\mathbb{E}[g(Y)|h(X)] = \mathbb{E}[g(Y)]$

Although $g(Y)$ and $h(X)$ are jointly continuous,

$\mathbb{E}[g(Y)|h(Y)]$ only takes the constant $\mathbb{E}[g(Y)]$,

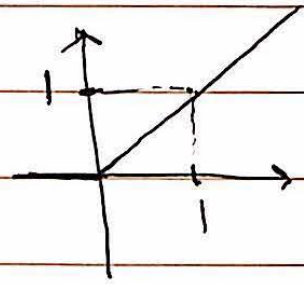
so it is a degenerate random variable and is discrete.

Mixed Random Variables

Example:

Consider the following function

$$g(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$



g is like a rectifier. If x is a positive voltage, $g(x) = x$.

If x is negative, $g(x) = 0$

Assume that $Y = g(X)$, where $X \sim U(-1,1)$. Let's Determine

$F_Y(y) = \mathbb{P}(Y \leq y)$

$X \in [-1,1] \Rightarrow Y \in$

$$\{\omega \mid Y(\omega) \leq y\} = \begin{cases} & y < 0 \\ & y \geq 1 \end{cases}$$

$$F_Y(y) = P(Y \leq y) = \begin{cases} & y < 0 \\ & y \geq 1 \end{cases}$$

But what is $F_Y(y)$ for $0 \leq y < 1$?

when $0 \leq y < 1$

$$F_Y(y) = P(Y \leq y) = P(\leq X \leq)$$

But X is uniform, so

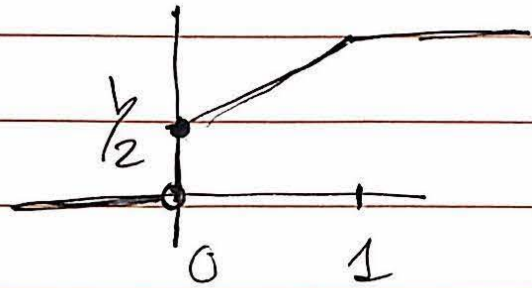
its pdf is $F_X(x) = x \quad 0 \leq x \leq 1$.

Then

$$P(\leq \alpha \leq) = F_X( ) - F_X( )$$

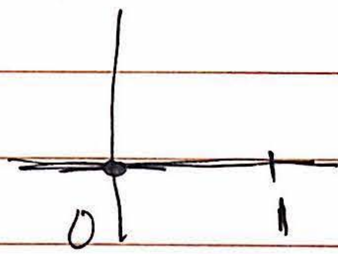
$=$

$$F_Y(y) = \begin{cases} 0 & y<0 \\ (y+1)/2 & 0 \le y < 1 \\ 1 & y \ge 1 \end{cases}$$

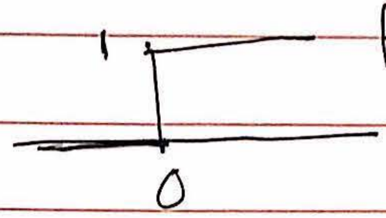


One can see F_Y to have a "step" singularity at $y=0$, and consider its derivative:

$f_Y(y) =$

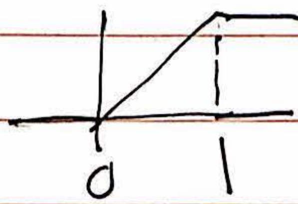


This random variable can be seen as a mixture of a discrete and a continuous r.v.:



F_{X_1}

$Y = X_1$ with Prob $1/2$



F_{X_2}

$Y = X_2$ with Prob $1/2$

A random variable that can be seen as a mixture of a discrete and continuous random variables has a density that contains impulse functions. We call such a random variable a "mixed random variable" and its density, a "generalized density."

A generalized density contains Dirac Delta Functions, (or impulse functions) characterized by the so-called sifting property:

$$\int_{-\infty}^{+\infty} h(t)\, \delta(t-t_0)\, dt = h(t_0)$$

The general form of a generalized density is

$$f_Y(y) = \tilde{f}_Y(y) + \sum_i \mathbb{P}(Y = y_i)\,\delta(y - y_i)$$

where y_i's are distinct and countable points at which

$F_Y(y)$ has jump discontinuities, due to ~~the~~ a discrete r.v. whose pmf is non-zero at y_i's, and $\tilde{f}_Y(y)$ is a continuous function without impulses, and $\forall y \neq y_i$

$$\frac{dF_Y}{dy} = f_{\tilde{Y}}(y)$$

The Expectation of any (measurable) function of Y, $g(Y)$ can be computed using the following mixed version of LOTUS

$$E[g(Y)] = \int_{-\infty}^{+\infty} g(y) f_Y(y)\, dy$$

$$= \int_{-\infty}^{+\infty} g(y) \tilde{f}_Y(y)\, dy + \sum_i g(y_i)\, \mathbb{P}(Y = y_i)$$

Exercise: Consider

$$f_Y(y) = \frac{5}{3} e^{-5y} + \frac{5}{12}\delta(y-1) + \frac{1}{4}\delta(y-5)$$

Compute $\mathbb{P}(1 < Y \leq 5)$,

$\mathbb{P}(Y = 1)$, and

$Var(Y)$.

Mixed Joint CDFS and PDFs

We learned that cdfs are fundamental concepts for any kind of random variable, therefore we can define

joint cdfs of discrete and continuous random variables.

Assume that N is a discrete and Y is a continuous r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$. Their joint cdf is defined as:

$$F_{Y,N}(y,n) = \mathbb{P}(Y \le y,\ N \le n)$$

One can even go further and define a generalized joint pdf for Y and N,

which contain "impulse_like" functions that have the property

$$\iint_{z=n} f(y,z) = 1.$$

(which is not possible for ordinary functions.

However, it is easier to study these cases by defining the following hybrid cdf-pmf function:

$$F_{Y,N}(y,n) = \mathbb{P}(Y \le y, N = n)$$

$$f_{Y,N}(y,n) = \frac{d}{dy}$$

and

$$F_{Y,N}(y,n) = \int_{-\infty}^{y}$$

Moreover the following hybrid conditional pdf and pmf can be defined

$f_{Y|N}(y|n) =$

$P_{N|Y}(n|y) =$

Consequently, the substitution law, the law of total probability, the Baye's Rule, and Conditional Expectation can be defined for hybrid pmf-cdfs:

The law of total probability

$P_N(n) =$

$f_Y(y) =$

The Baye's Rule can be formulated as

$f_{Y|N}(y|n) =$

$p_{N|Y}(n|y) =$

Exercise: Formulate the substitution law, conditional expectation, and the law of total expectation for $F_{Y,N}(y,n)$.

Exercise: Let X, Y be jointly continuous random variables with joint pdf

$$f_{X,Y}(x,y) = \begin{cases} cx+1 & x, y \geq 0,\ x+y<1 \\ 0 & \text{otherwise} \end{cases}$$

1- Find c

2- Find the marginal pdfs $f_X(x)$ and $f_Y(y)$

3- Find $\mathbb{P}(Y < 2X^2)$

4- Find $\mathbb{E}[Y \mid X > 1/2]$

Exercise: Let X, Y be jointly continuous r.v.'s with joint pdf

$$f_{X,Y}(x,y) = \begin{cases} 6e^{-(2x+3y)} & x, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

1. Are X and Y independent?

2. Find $\mathbb{E}[Y \mid X > 2]$

3. Find $\mathbb{P}(X > Y)$.

Types of Random Variables

There are three fundamentally different kinds of random variables. The Lebesgue Decomposition Theorem states that any

probability law $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ can be uniquely decomposed into a sum of these three types of measures:

- Discrete
- Continuous, and

- Singular

In other words, there are three pure types of random variables: discrete, continuous, and singular. It is also possible to mix these three types, to have overall seven types of random variables.

Of the three pure types, only discrete and continuous random variables are of practical interest. That is why we spent a lot of time on those random variables. For the sake of completeness, here we briefly discuss singular random variables.

Singular random variables are strange. In some sense, they resemble both discrete and continuous random variables: Singular Random

Variables take values with probability 1 on an uncountable set of Lebesgue measure zero!

Remark: We never defined the

meaning of a Lebesgue measure. However, we intuitively observed that it represents a generalized concept of length (in one dimension) and area or volume in higher dimensions. Countable sets do not produce any length, area, or volume. So, they are sets of Lebesgue Measure Zero. But there are even

uncountable sets of Lebesgue measure zero!

Let's see the construction of such a set, called the Cantor Set, C, and building a CDF on it.

The Cantor Set is defined by removing $(1/3, 2/3)$ from $[0,1]$ and then removing the middle third of each interval that remains.

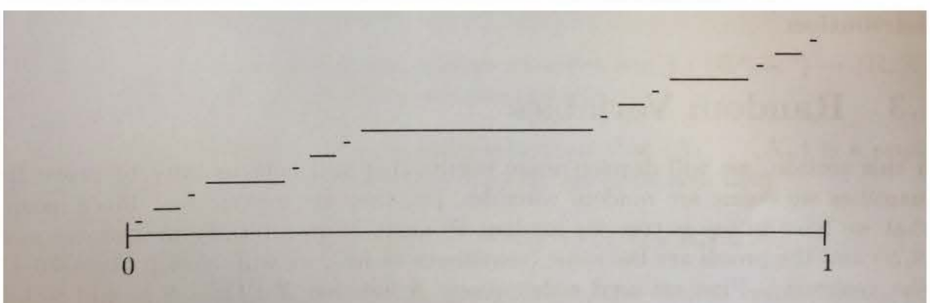
We define an associated distribution function by setting $F(x) = 0$ for $x \leq 0$, $F(x) = 1$ for $x \geq 1$, $F(x) = 1/2$ for $x \in [1/3, 2/3]$,

$F(x) = 1/4$ for $x \in [1/9, 2/9]$, $F(x) = 3/4$ for $x \in [7/9, 8/9]$ ...

The CDF constructed by this procedure is called the Cantor Function (sometimes called the Devil's Staircase). The Cantor Function is continuous everywhere, since all singletons have zero probability under this distribution. Also, the derivative is zero wherever it exists, and the derivative does not exist at points in the Cantor Set C

(The CDF only increases at these Cantor points, but does

So without a well-defined derivative, or any jump discontinuities!



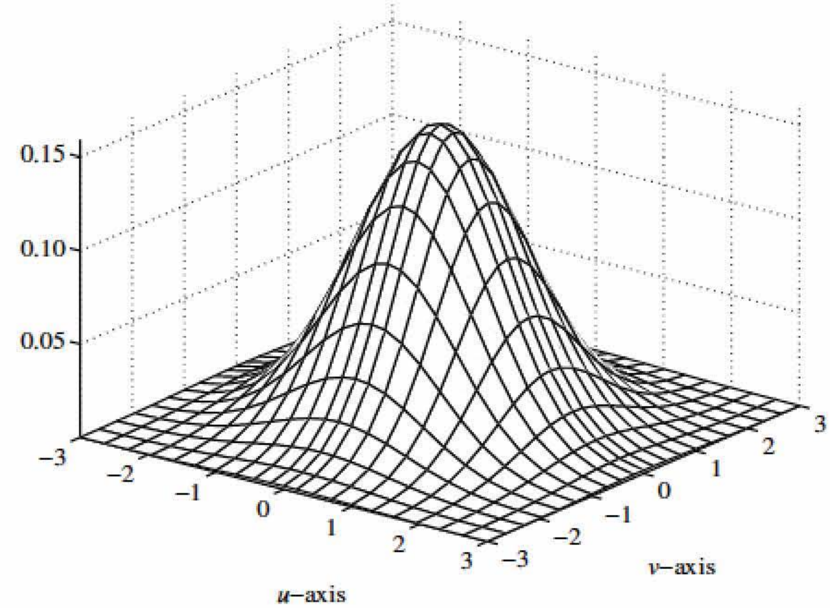
Cantor distribution function

The Bivariate Normal Distribution

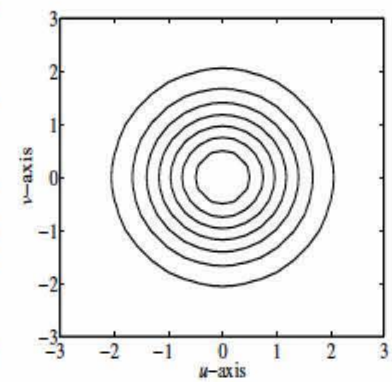
Assume that U and V are jointly continuous r.v.'s and $\rho \in (-1,1)$. U and V are called to have a standard bivariate normal density, iff their pdf is:

$$f_{UV}(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right)$$

This function is shown below for $\rho = 0$ on the u and v axes:



From the formula, one can see that for $\rho = 0$, this function is circulary symmetric, i.e. on a circle of radius r, its value is:



$u^2 + v^2 = r^2$, $\rho \geq 0$

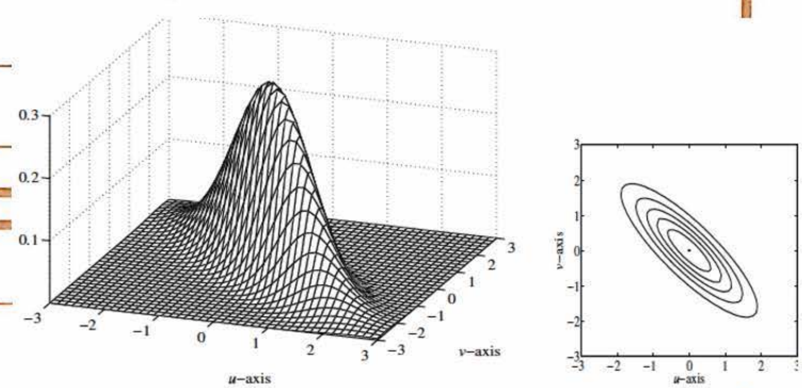
$\Rightarrow \quad f_{U,V}(u,v) =$

which $\dfrac{\text{does}}{\text{doesn't}}$ depend

on the particular value of u, v. Contours

of $f_{U,V}$ is shown above.

A plot of f_{UV} for $\rho = -.85$ is shown below. Now f_{UV} is constant on ellipses instead of circles. The axes of the ellipse are not parallel to the coordinate axes. The major axes of these ellipses and the density are concentrated along $v = -u$ and when $\rho \to -1$, this concentration becomes a "delta" function.



Also for $\rho = 0$, the joint pdf decomposes into two standard normal r.v.'s

$$f_{UV}(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right)$$

Lemma: Let U and V be two jointly continuous r.v.'s defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with joint pdf

$$f_{U,V} = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right)$$

(1) f is indeed a pdf.

(2) The marginal density of U, V is standard normal

(3) $\rho(U,V) = \rho$. Moreover U, V are independent iff $\rho = 0$.

(4) $f_{V|U}(v|u) \sim \mathcal{N}(\rho u, 1-\rho^2)$

Proof:

2) completing the squares

$u^2 - 2\rho uv + v^2 =$

$$f_V(v) = \int_{-\infty}^{+\infty} f_{U,V}(u,v)\, du$$

$=$

Because of symmetry, f_V is also that of $N(0,1)$.

1)

$$\int_{-\infty}^{+\infty} \underbrace{\int_{-\infty}^{+\infty} f_{U,V}(u,v)\, du}_{f_V} dv =$$

3) $\mathrm{Cov}(U,V) = E[UV] - \underbrace{E[U]}_{0}\underbrace{E[V]}_{0}$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} uv\, f_{U,V}(u,v)\, du\, dv$$

Let's first calculate

$\int_{-\infty}^{+\infty} u f_{U,V}(u,v)\, du =$

Next,

$$E[UV] = \int_{-\infty}^{+\infty} \rho v\, v\, f_V(v)\, dv$$

$=$

$=$

Therefore $Cov(U, V) = \rho$

$$\rho(U, V) = \frac{Cov(U, V)}{\sqrt{Var(U)}\sqrt{Var(V)}} =$$

Moreover, we know that if

U, V are independent, $\rho(U, V) = 0$

$\Rightarrow \rho = 0$

The reverse is not necessarily true,

but it is true for jointly normal

r.v.'s because

$\rho = 0 \Rightarrow f_{U,V}(u,v) =$

(4) $f_{V|U}(v|u) = \dfrac{f_{U,V}(u,v)}{f_U(u)}$

$=$

The General Bivariate Normal Density

$$f(u,v)=\frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-\rho^2}}\exp\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(u-\mu_u)^2}{\sigma_u^2}-2\rho\left(\frac{u-\mu_u}{\sigma_u}\right)\left(\frac{v-\mu_v}{\sigma_v}\right)+\frac{(v-\mu_v)^2}{\sigma_v^2}\right]\right)$$

Exercise: Verify that

$\mathbb{E}[U]=\mu_u \qquad Var(U)=\sigma_u^2$

$\mathbb{E}[V]=\mu_v \qquad Var(V)=\sigma_v^2$

$\rho(U,V)=\rho$