ECON2125/4021/8013

Lecture 13

John Stachurski

Semester 1, 2015

Announcements

The lecture 12 PDF was shortened after the lecture

- Last part of lecture was removed ex post (pages 42 onwards)
    - The discussion of densities and pmfs
    - Replaced with a better treatment in this lecture (lecture 13)
- See GitHub for updated lecture 12 PDF (pages 1–41 only)
- For hardcopy versions, just discard pages 42 onwards

Discussion of exam tomorrow

Densities and Probability Mass Functions

Recall that the distribution of random variable X is the function

$$F(x) := \mathbb{P}\{X \leq x\} \qquad (x \in \mathbb{R})$$

- Contains useful information about X and $\mathbb{P}$
- Always a cdf

But cdfs are not always intuitive

- convey information about probability mass through slope
- harder to read than height, say
- and how do we integrate using these things?

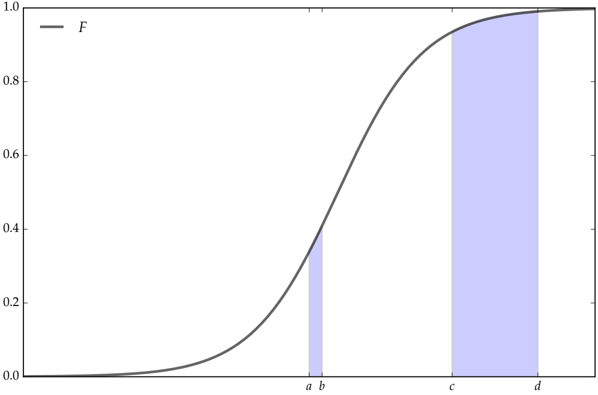


Figure : Which interval has more probability mass?

There are two special cases where distributions can be represented by simpler objects:

1. Discrete random variables
2. Continuous random variables

Not every RV fits into one of these categories

But when it does things are simpler

Remarks on terminology:

- Discrete RVs have distributions that increase only through discrete jumps
- Continuous RVs have continuous distributions

The Density Case

Let's start with the case of continuous random variables

A **density function** on $\mathbb{R}$ is a function $p\colon \mathbb{R} \to \mathbb{R}$ such that

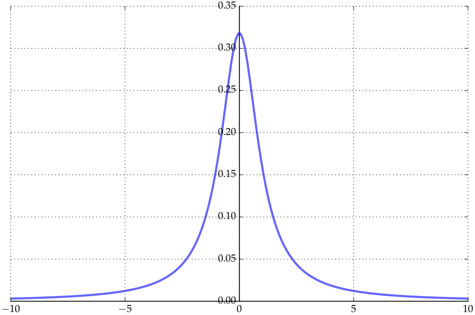
1. $p(x) \geq 0$ for all $x \in \mathbb{R}$
2. p integrates to one — that is,

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

Example. The function

$$p(x) = 1/(\pi + \pi x^2)$$

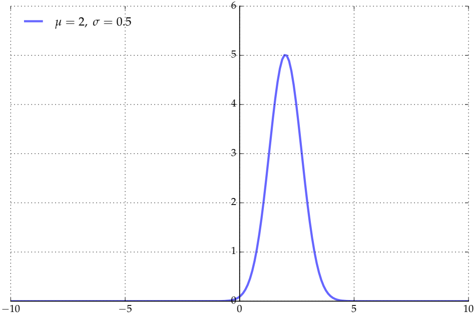
is a density called the **Cauchy density**



Example. For any $\mu \in \mathbb{R}$ and $\sigma > 0$,

$$p(x) = (2\pi\sigma^2)^{-1/2} \exp(-(x-\mu)^2/(2\sigma^2))$$

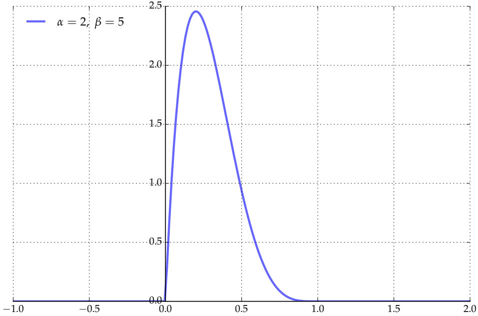
is a density called the **normal density**, and written $N(\mu, \sigma^2)$



Example. For any $\alpha, \beta > 0$,

$$p(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 r^{\alpha-1}(1-r)^{\beta-1}\,dr} & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

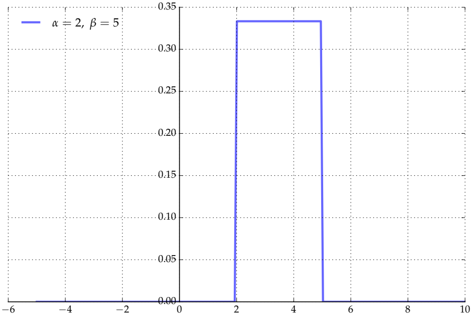
is a density called the **beta density**, and written $B(\alpha, \beta)$



Example. For any $\alpha < \beta$,

$$p(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } x \in (\alpha, \beta) \\ 0 & \text{otherwise} \end{cases}$$

is a density called the **uniform density**, and written $U(\alpha, \beta)$

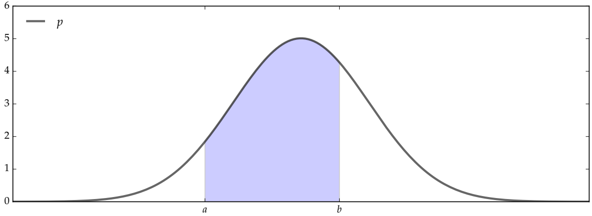


A random variable X is said to **have density** p if

1. p is a density
2. X satisfies

$$\mathbb{P}\{a < X \leq b\} = \int_a^b p(x)dx$$

for any $a < b$

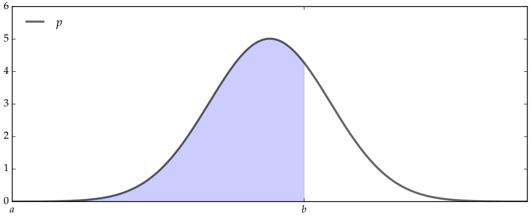


Here we allow $a = -\infty$ and $b = +\infty$

For example

$$\mathbb{P}\{X \leq b\} = \mathbb{P}\{-\infty < X \leq b\}$$

$$= \int_{-\infty}^{b} p(x)dx$$



Connection to the Distribution F

Follows directly from last slide that the distribution F of X satisfies

$$F(x) = \int_{-\infty}^{x} p(s)ds \quad \text{for all} \quad x \in \mathbb{R}$$

Facts Let F be the distribution of X

1. If X has a density p then F is continuous
2. If p is continuous at x then F is differentiable at x and

$$F'(x) = p(x)$$

Proof: See the Fundamental Theorem of Calculus

Example. Recall the Cauchy cdf

$$F(x) = \arctan(x)/\pi + 1/2$$

and the Cauchy density

$$p(x) = \frac{1}{\pi(1+x^2)}$$

The density is continuous everywhere, and, since

$$\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$$

we have

$$F'(x) = \frac{1}{\pi(1+x^2)} = p(x)$$

Example. The $U(\alpha, \beta)$ cdf is continuous but not differentiable at α and β

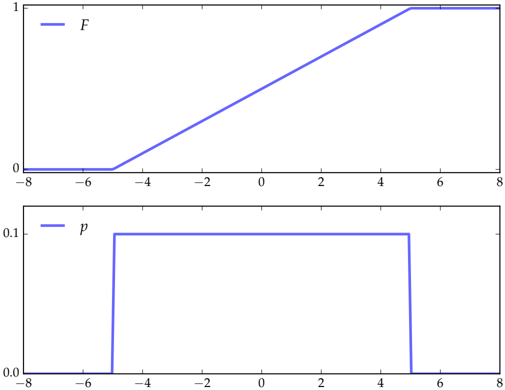


Figure : Uniform density and cdf with $\alpha = -5$ and $\beta = 5$

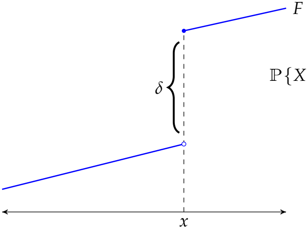
Fact. If X has any density then

$$\mathbb{P}\{X = x\} = 0 \quad \text{for every} \quad x \in \mathbb{R}$$

Proof: $\exists\, x \in \mathbb{R}$ with $\mathbb{P}\{X = x\} > 0$ contradicts continuity of F

Indeed, if $\mathbb{P}\{X = x\} = \delta > 0$, then, $\forall\, \epsilon > 0$,

$$F(x) - F(x - \epsilon) = \mathbb{P}\{x - \epsilon < X \leq x\} \geq \delta$$



$\mathbb{P}\{X = x\} = \delta \implies F$ jumps by δ at x

Fact. If X is a random variable with density p and $a \le b$, then all of the following are true

$$\mathbb{P}\{a < X \le b\} = \int_a^b p(x)dx$$

$$\mathbb{P}\{a \le X < b\} = \int_a^b p(x)dx$$

$$\mathbb{P}\{a \le X \le b\} = \int_a^b p(x)dx$$

$$\mathbb{P}\{a < X < b\} = \int_a^b p(x)dx$$

Let's just check the case

$$\mathbb{P}\{a < X < b\} = \int_a^b p(x)dx$$

Proof: We have

$$\{a < X \le b\} = \{a < X < b\} \cup \{X = b\}$$

$$\therefore \quad \mathbb{P}\{a < X \le b\} = \mathbb{P}\{a < X < b\} + \mathbb{P}\{X = b\}$$

$$\therefore \quad \mathbb{P}\{a < X \le b\} = \mathbb{P}\{a < X < b\}$$

By definition, the LHS is $\int_a^b p(x)dx$

The Discrete Case

A **probability mass function** (**pmf**) is a sequence $\{p_k\}$ such that

1. $p_k \geq 0$ for all $k \geq 0$
2. p_k sums to unity:

$$\sum_{k \geq 0} p_k = 1 \qquad \text{(finite or infinite sum)}$$

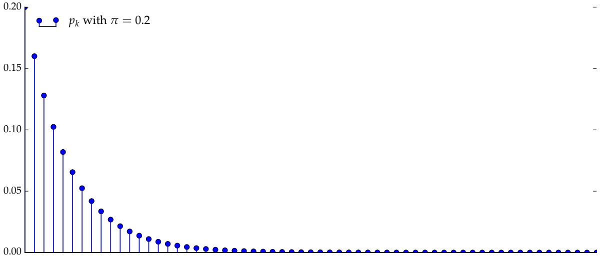
Here $\{p_k\}$ is countable — that is,

- finite, as in $\{p_0, \ldots, p_K\}$ or
- countably infinite, as in $\{p_0, p_1, \ldots\}$

Example. Given $\pi \in (0,1)$,

$$p_k = (1-\pi)^k \pi \qquad k = 0, 1, \ldots$$

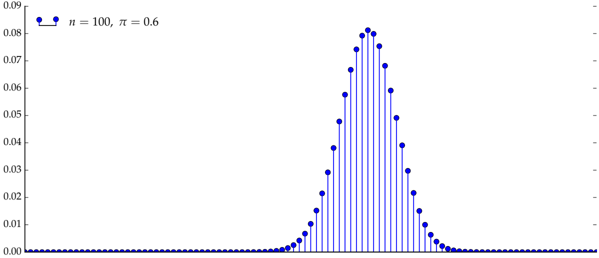
is a pmf called the **geometric pmf**



Example. Given $n \in \mathbb{N}$ and $\pi \in (0,1)$,

$$p_k = \binom{n}{k} \pi^k (1-\pi)^{n-k}, \qquad k = 0, \ldots, n$$

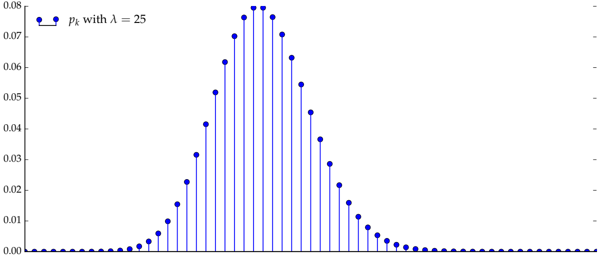
is a pmf called the **binomial pmf**, and written $B(n, \pi)$



Example. Given $\lambda > 0$,

$$p_k = \frac{\exp(-\lambda)\lambda^k}{k!} \qquad k = 0, 1, \ldots$$

is a pmf called the **poisson pmf**, and written $P(\lambda)$



A random variable X with range $\{x_k\}$ is said to **have pmf** $\{p_k\}$ if

1. $\{p_k\}$ is a pmf
2. $\mathbb{P}\{X = x_k\} = p_k \quad$ for all $k \geq 0$

Notes:

- $|\operatorname{rng}(X)| = |\{p_k\}| \quad$ (same cardinality)
- when $\{x_k\}$ not made explicit you can assume

$$x_k = k \quad \text{for all} \quad k$$

Connection to the Distribution F

Let X be a RV with range $\{x_k\}$, pmf $\{p_k\}$

Fact. The distribution of X satisfies

$$F(x) = \sum_{k \geq 0} \mathbb{1}\{x_k \leq x\} p_k \qquad (x \in \mathbb{R})$$

Proof for finite case:

$$\mathbb{P}\{X \leq x\} = \sum_{k=0}^{K} \mathbb{P}\{X \leq x \mid X = x_k\} \mathbb{P}\{X = x_k\}$$

$$= \sum_{k=0}^{K} \mathbb{1}\{x_k \leq x\} p_k$$

Visually, F is a step function with jump p_k at x_k

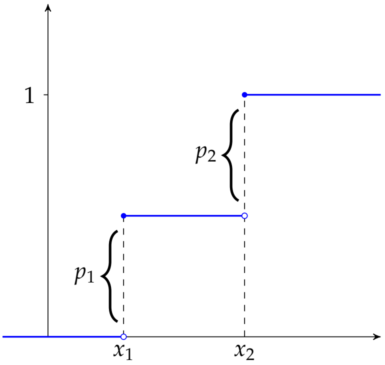


Figure : $F(x) = \mathbb{1}\{x_1 \leq x\}p_1 + \mathbb{1}\{x_2 \leq x\}p_2$

Expectations from Distributions

Let $h\colon \mathbb{R} \to \mathbb{R}$

We often want to calculate an expectation such as $\mathbb{E}\left[h(X)\right]$

Examples.

- If $h(x) = x$ then $\mathbb{E}\left[h(X)\right]$ is the expectation of X
  - Sometimes called the **mean**
- If $h(x) = x^2$ then $\mathbb{E}\left[h(X)\right]$ is the second moment
- If μ is the mean of X and $h(x) = (x - \mu)^2$, then $\mathbb{E}\left[h(X)\right]$ is the variance of X

We can use formal definition of expectations to obtain $\mathbb{E}[h(X)]$

But often that's hard work

On the other hand, if

1. $X \sim F$
2. F has nice properties

this can help us compute the expectation

This is true particularly when F is generated by a density or pmf

The details follow

Fact. If X has pmf $\{p_k\}$ with range $\{x_k\}$ then

$$\mathbb{E}\left[h(X)\right] = \sum_{k \geq 0} h(x_k) p_k$$

whenever the RHS is finite

Example. A geometric RV with parameter $\pi \in (0, 1)$ is a discrete RV with

$$\mathbb{P}\{X = k\} = (1 - \pi)^k \pi \qquad (k = 0, 1, \ldots)$$

Setting $x_k = k$ and $h(x) = x$, we have

$$\mathbb{E}\left[X\right] = \sum_{k=0}^{\infty} k(1-\pi)^k \pi = \pi \sum_{k=0}^{\infty} k(1-\pi)^k = \pi \frac{1-\pi}{\pi^2} = \frac{1-\pi}{\pi}$$

Fact. If X has density p, then

$$\mathbb{E}\left[h(X)\right] = \int_{-\infty}^{\infty} h(x)p(x)dx$$

whenever the RHS is finite

Example. If $X \sim N(0,1)$ then

$$\mathbb{E}\left[X\right] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = 0$$

$$\mathrm{var}[X] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = 1$$

(You won't be asked to solve these integrals in the exams)

Unifying Notation

Convenient notation to unify:

$$\text{If } X \sim F, \text{ we write } \mathbb{E}\left[h(X)\right] = \int h(x)F(dx)$$

Meaning:

- In the density case

$$\int h(x)F(dx) := \int_{-\infty}^{\infty} h(x)p(x)dx$$

- In the discrete case

$$\int h(x)F(dx) := \sum_{k \geq 0} h(x_k)p_k$$

Neither Density nor PMF

Some cdfs fit neither the density nor the discrete case

- mix jumps and smooth increases

For this case we can still write

$$\mathbb{E}\left[h(X)\right] = \int h(x) F(dx)$$

where the RHS is the "Lebesgue-Stieltjes integral" with respect to F

- A bit too advanced for this course
- We will stick mainly to the density or pmf cases

Joint Distributions

Consider N random variables $X_1, \ldots, X_N$, where $X_n \sim F_n$

F_n tells us about properties of X_n viewed as a single entity

How about the relationships between the variables $X_1, \ldots, X_N$?

To quantify, we define the **joint distribution** of $X_1, \ldots, X_N$ to be

$$F(x_1, \ldots, x_N) := \mathbb{P}\{X_1 \leq x_1, \ldots, X_N \leq x_N\}$$

In this setting, F_n sometimes called the **marginal distribution**

Random Vectors

We can also view $X_1, \ldots, X_N$ collectively as a **random vector** $\mathbf{X}$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}$$

Sometimes $\mathbf{X}$ will be a row vector:

$$\mathbf{X} = \begin{pmatrix} X_1 & \cdots & X_N \end{pmatrix}$$

The distribution of $\mathbf{X}$ is just the joint distribution of $X_1, \ldots, X_N$

$$F(\mathbf{x}) = \mathbb{P}\{\mathbf{X} \leq \mathbf{x}\} := \mathbb{P}\{X_1 \leq x_1, \ldots, X_N \leq x_N\}$$

For random vector $\mathbf{X}$, the expectation is defined pointwise

- Row case

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] & \cdots & \mathbb{E}[X_N] \end{pmatrix}$$

- Column case

$$\mathbb{E}[\mathbf{X}] := \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_N] \end{pmatrix}$$

Unless otherwise specified, we treat $\mathbf{X}$ and $\mathbb{E}[\mathbf{X}]$ as column vectors

Linearity carries over to the vector setting

For example, if

1. $\mathbf{X}$ and $\mathbf{Y}$ are random vectors
2. $\mathbf{A}$ and $\mathbf{B}$ are conformable matrices

then all the following are true

- $\mathbb{E}[\mathbf{AX}] = \mathbf{A}\mathbb{E}[\mathbf{X}]$
- $\mathbb{E}[\mathbf{XA}] = \mathbb{E}[\mathbf{X}]\mathbf{A}$
- $\mathbb{E}[\mathbf{AXB}] = \mathbf{A}\mathbb{E}[\mathbf{X}]\mathbf{B}$
- $\mathbb{E}[\mathbf{X}+\mathbf{Y}] = \mathbb{E}[\mathbf{X}] + \mathbb{E}[\mathbf{Y}]$

Proofs: Just break it down and check element by element

The **variance-covariance matrix** of $\mathbf{X}$

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

Letting $\mu_i := \mathbb{E}\left[X_i\right]$ and expanding this out we get

$$\begin{pmatrix} \mathbb{E}[(X_1 - \mu_1)(X_1 - \mu_1)] & \cdots & \mathbb{E}[(X_1 - \mu_1)(X_N - \mu_N)] \\ \mathbb{E}[(X_2 - \mu_2)(X_1 - \mu_1)] & \cdots & \mathbb{E}[(X_2 - \mu_2)(X_N - \mu_N)] \\ \vdots & \vdots & \vdots \\ \mathbb{E}[(X_N - \mu_N)(X_1 - \mu_1)] & \cdots & \mathbb{E}[(X_N - \mu_N)(X_N - \mu_N)] \end{pmatrix}$$

- The j,k-th element is the covariance between X_j and X_k
- The principle diagonal contains the variance of each X_n

Fact. For any random vector $\mathbf{X}$ and conformable nonrandom $\mathbf{a}$ and $\mathbf{B}$, we have $\mathrm{var}[\mathbf{a} + \mathbf{B}\mathbf{X}] = \mathbf{B}\,\mathrm{var}[\mathbf{X}]\mathbf{B}'$

Proof: Letting $\boldsymbol{\mu} := \mathbb{E}[\mathbf{X}]$ and using linearity of $\mathbb{E}$ we have

$$\begin{aligned}
\mathrm{var}[\mathbf{a} + \mathbf{B}\mathbf{X}] &= \mathbb{E}\left[(\mathbf{a} + \mathbf{B}\mathbf{X} - \mathbf{a} - \mathbf{B}\boldsymbol{\mu})(\mathbf{a} + \mathbf{B}\mathbf{X} - \mathbf{a} - \mathbf{B}\boldsymbol{\mu})'\right] \\
&= \mathbb{E}\left[\mathbf{B}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{B}(\mathbf{X} - \boldsymbol{\mu}))'\right] \\
&= \mathbb{E}\left[\mathbf{B}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\mathbf{B}'\right] \\
&= \mathbf{B}\mathbb{E}\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\right]\mathbf{B}' \\
&= \mathbf{B}\,\mathrm{var}[\mathbf{X}]\mathbf{B}'
\end{aligned}$$

Fact. $\mathrm{var}[\mathbf{X}]$ is always symmetric and nonnegative definite

The Density Case

As for scalar case, some random vectors have densities or pmfs

Make life easier when they exist

For the vector case we will focus on densities, skip pmfs

A **density function** on $\mathbb{R}^N$ is a function $p \colon \mathbb{R}^N \to \mathbb{R}$ such that

1. $p(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^N$
2. p integrates to one — that is,

$$\int p(\mathbf{x}) d\mathbf{x} := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \ldots, x_N) dx_1 \cdots dx_N = 1$$

A random N-vector $\mathbf{X}$ is said to **have density p** if

1. p is a density on $\mathbb{R}^N$
2. for any $a_1 < b_1, \ldots, a_N < b_N$, $\mathbf{X}$ satisfies

$$\mathbb{P}\{a_1 < X_1 \leq b_1, \ldots, a_N < X_N \leq b_N\}$$

$$= \int_{a_N}^{b_N} \cdots \int_{a_1}^{b_1} p(x_1, \ldots, x_N) dx_1 \cdots dx_N$$

In particular,

$$F(\mathbf{x}) = \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_1} p(z_1, \ldots, z_N) dz_1 \cdots dz_N$$

Example. The **multivariate normal density** on $\mathbb{R}^N$ is the density

$$p(\mathbf{x}) = (2\pi)^{-N/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Here

- $\boldsymbol{\mu}$ is any $N \times 1$ vector
- $\boldsymbol{\Sigma}$ is a symmetric, positive definite $N \times N$ matrix

In symbols, we represent this distribution by $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

We say that $\mathbf{X}$ is **standard normal** if $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$

Fact. If $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu} \quad \text{and} \quad \text{var}[\mathbf{X}] = \boldsymbol{\Sigma}$$

Example. Continuing the previous example, if $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

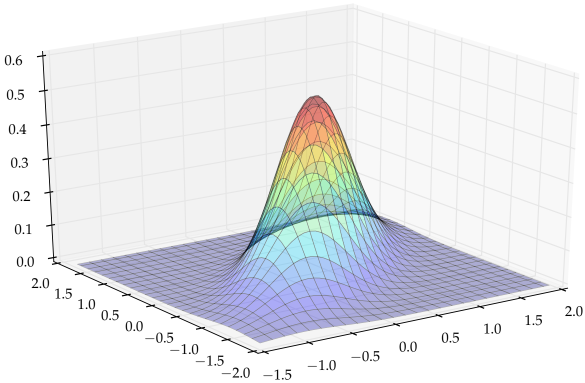
and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}$$

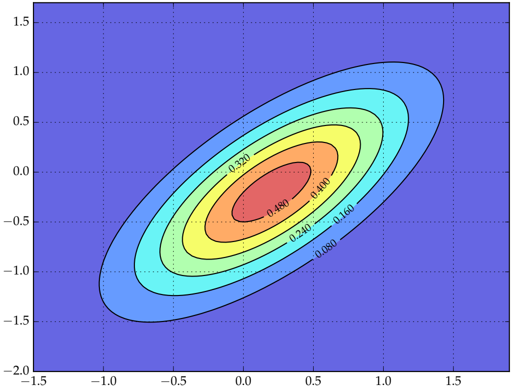
then

- $\mathbb{E}[X_i] = \mu_i$ for $i = 1, 2$
- $\text{var}[X_i] = \sigma_i^2$ for $i = 1, 2$
- $\text{cov}[X_1, X_2] = \rho$
- $\text{corr}[X_1, X_2] = \rho / (\sigma_1 \sigma_2)$

Example. $\mu_1 = 0.2,\ \mu_2 = -0.2,\ \rho = 0.3,\ \sigma_1^2 = 0.4,\ \sigma_2^2 = 0.45$



Example. $\mu_1 = 0.2,\ \mu_2 = -0.2,\ \rho = 0.3,\ \sigma_1^2 = 0.4,\ \sigma_2^2 = 0.45$



Linearity and Normality

Fact. If $\mathbf{X}$ is normal then so is $\mathbf{a} + \mathbf{B}\mathbf{X}$ for any conformable nonrandom $\mathbf{a}$ and $\mathbf{B}$

Example. Consider $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}$$

If $Y = \alpha X_1 + \beta X_2$ then Y is normal by above fact

Ex. Show that its mean and variance are

- $\mathbb{E}[Y] = \alpha\mu_1 + \beta\mu_2$
- $\text{var}[Y] = \alpha^2\sigma_1^2 + \beta^2\sigma_2^2 + 2\alpha\beta\rho$

Fact. If $\mathbf{X} = (X_1, \ldots, X_N)$ is multivariate normal then each X_i is univariate normal

Proof: Follows from previous fact and $X_i = \mathbf{e}_i' \mathbf{X}$

Example. Consider $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}$$

We already know that $\mathbb{E}[X_i] = \mu_i$ and $\text{var}[X_i] = \sigma_i^2$

Since X_i is normal we have

$$X_i \sim N(\mu_i, \sigma_i^2) \text{ for } i = 1, 2$$

From Joint to Marginal

Marginals can be recovered from joint distributions by integrating

Example. If X_1, X_2 have joint density p on $\mathbb{R}^2$, then X_2 has marginal density

$$q(x_2) := \int_{-\infty}^{\infty} p(x_1, x_2) dx_1$$

True because

$$\mathbb{P}\{a < X_2 \leq b\} = \mathbb{P}\{-\infty < X_1 < \infty,\ a < X_2 \leq b\}$$

$$= \int_a^b \left[\int_{-\infty}^{\infty} p(x_1, x_2) dx_1 \right] dx_2$$

$$= \int_a^b q(x_2) dx_2$$

Conditional Distributions

The **conditional density** of Y given $X = x$ is defined by

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

Here and below we use loose but common notation:

- $p(x,y)$ is the joint density of (X,Y)
- $p(x)$ is the marginal density of X
- $p(y)$ is the marginal density of Y
- etc.

Law of Total Probability, Density Case

Fact. If (X, Y) is a random vector with density p, then

$$p(y) = \int_{-\infty}^{\infty} p(y \mid x) p(x) dx \qquad (y \in \mathbb{R}) \tag{1}$$

Proof: Fix $y \in \mathbb{R}$

Integrating the joint to get the marginal, we have

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

Now combine this with $p(y \mid x) = p(x, y) / p(x)$ to get (1)