Math-UA.233: Theory of Probability
Lecture 15

Tim Austin

From last time: normal RVs

A RV X is $\mathrm{N}(\mu, \sigma^2)$ if it is continuous and has PDF

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad \text{for } x \in \mathbb{R}.}$$

In particular, if X is $\mathrm{N}(0, 1)$:

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{for } x \in \mathbb{R}.}$$

Normal RVs are all related to each other by transformations:

Proposition (Ross p188)

If X is $N(\mu, \sigma^2)$ then $aX + b$ is $N(a\mu + b, a^2\sigma^2)$ for any $a > 0$ and any b.

PROOF: Use formula for the PDF of a transformed RV.

In particular,

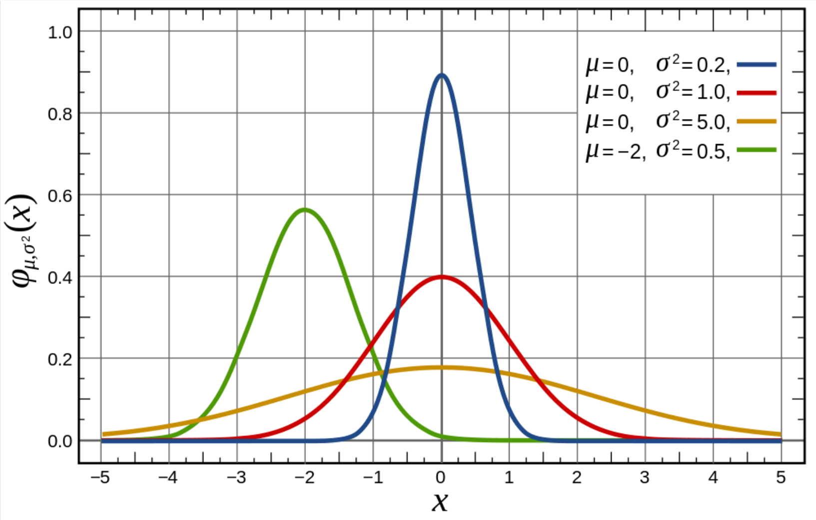
$$\text{if } X \text{ is } N(\mu, \sigma^2), \text{ then } Z = \frac{X - \mu}{\sigma} \text{ is standard normal.}$$

This Z is called the **standard version** of X. Problems about X can be translated into problems about Z: this process is called **standardizing**.

First application of standardizing:

Proposition (Ross E.g. 5.4a)

If X is $N(\mu, \sigma^2)$ then $E[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$.



1.0
0.8
0.6
0.4
0.2
0.0
$\varphi_{\mu,\sigma^2}(x)$
$\mu=0, \quad \sigma^2=0.2,$
$\mu=0, \quad \sigma^2=1.0,$
$\mu=0, \quad \sigma^2=5.0,$
$\mu=-2, \quad \sigma^2=0.5,$
−5 −4 −3 −2 −1 0 1 2 3 4 5
x

The standard normal CDF

Many problems boil down to finding some values of the standard normal CDF: that is, the function

$$\boxed{\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy}.$$

This important function is always denoted by Φ, because *the above integral has no simple closed formula*! (Deep fact from outside this course.)

Instead, we often use tables of values of Φ, or some scientific calculators have it as a separate function (alongside sin, exp etc.).

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of th

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992

Many tables give $\Phi(a)$ only for $a \geqslant 0$. Other values can be obtained from the important equation

$$\boxed{\Phi(-a) = 1 - \Phi(a)}$$

(Ross p190).

If X is $\mathrm{N}(\mu, \sigma^2)$, then we can obtain its CDF in terms of Φ by standardizing:

$$\boxed{F_X(a) = P(X \leqslant a) = P\left(\frac{X-\mu}{\sigma} \leqslant \frac{a-\mu}{\sigma}\right) = \Phi\Big(\frac{a-\mu}{\sigma}\Big).}$$

Example (Ross E.g. 5.4b)

If X is $N(3, 9)$, find (a) $P(2 < X < 5)$, (b) $P(X > 0)$ and (c) $P(|X - 3| > 6)$.

Example (Part of Ross E.g. 5.4c)

If X is $N(\mu, \sigma^2)$, find $P(X > \mu + \sigma)$.

Exponential RVs (Ross Sec 5.5)

Many important continuous RVs occur when modeling the **waiting time** before a random 'event' (e.g. earthquake, failure of a device, death of an individual) (BEWARE: not our usual, technical use of the term 'event').

Different RVs are needed depending on what situation is being modeled.

We have studied a special class of such situations before:

Events occur at random moments. Fix a 'rate' parameter $\lambda > 0$, and assume that:

- If *h* is very short, then the probability of an 'event' during an interval of length *h* is approximately λh (up to an error which is much smaller than *h*).
- If *h* is very short, then the probability of two or more 'event's during an interval of length *h* is negligibly small.
- For any time-intervals which don't overlap, the numbers of 'event's that occur in these time intervals are independent.

Then we have seen previously (Lecture 11) that

> The number of events that occur in a time-interval of length *t* is Poi(λt).

Now let's ask a different question about the same class of situations:

> *Let T be the time from now until the next 'event'. What is the distribution of the RV T?*

SOLUTION: Obviously $T \geqslant 0$. We can derive the CDF of T as follows: for any threshold $t > 0$, we have

$$\begin{aligned} P(T > t) &= P(\text{no events during inteval } [0, t]) \\ &= P(\# \text{ events during interval } [0, t] \text{ equals } 0) \\ &= P(N = 0) \quad \text{where } N \text{ is } \mathrm{Poi}(\lambda t) \\ &= e^{-\lambda t} \\ \therefore \quad P(T \leqslant t) &= 1 - P(T > t) = 1 - e^{-\lambda t}. \end{aligned}$$

Since we have

$$P(T \leqslant t) = 1 - e^{-\lambda t} = \int_0^t \lambda e^{-\lambda s}\, ds,$$

we have shown that

T is a continuous RV with PDF

$$f(s) = \begin{cases} 0 & \text{if } s < 0 \\ \lambda e^{-\lambda s} & \text{if } s \geqslant 0. \end{cases}$$

(You can also regard this RV T as a continuous cousin of a geometric RV, which describes the waiting time for success in a discrete sequence of trials. See Pishro-Nik, Solved Problem 4.2.6 # 2 for more on this.)

As previously, we turn this important calculation into a definition.

Definition

*Let $\lambda > 0$ be a parameter. A RV X is **exponential with parameter** λ, or just **Exp**(λ), if it is continuous with PDF*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geqslant 0 \\ 0 & \text{for } x < 0 \end{cases}$$

So exponential RVs should be used to model waiting times in all the same situations where we can use Poisson RVs to model ‘number of events in a given time interval’.

Reminder of some examples:

- time until the next earthquake in California,
- time until a wrong number is received at a telephone exchange,
- time until a radioactive atom decays.

Non-example:

- Time until the next uptown C train pulls into West 4th Street station. Why isn’t this an example? Different time intervals not independent!

Basic facts about exponential RVs:

Proposition (Ross p198)

If X is Exp(λ) then its CDF is

$$F(a) = 1 - e^{-\lambda a} \quad \text{for } a \geqslant 0.$$

Proposition (Ross E.g. 5.5a)

If X is Exp(λ) then (a) $E[X] = 1/\lambda$ and (b) $\mathrm{Var}(X) = 1/\lambda^2$.

Example (Modified from Ross E.g. 5.5b)

You are a birdwatcher looking for a chaffinch. The time until your next sighting is an Exp$(1/10)$ *RV. Find the probability that you have to wait*

(a) *more than 10 minutes.*
(b) *between 10 and 20 minutes.*

(Pause for thought: why does this e.g. fit the general situation we talked about?)

Exponential RVs have a very important property.

Proposition (The memoryless property; Ross p199)

If X is $Exp(\lambda)$ for some λ, and $s, t \geqslant 0$, then

$$\boxed{P(X > s + t \,|\, X > t) = P(X > s).}$$

So given that we've already waited t hours, the probability that we have to wait an additional s hours is the same as the unconditioned probability of waiting s hours if we start over.

(And in fact, exponential RVs are the *only* non-negative RVs which have this property: see Ross pp199-200 for the proof.)

The memoryless property is extremely important, because it makes certain calculations possible even when you don't seem to have complete information.

Example (Modified from Ross E.g. 5.5d)

The number of miles that a car can run before getting a puncture is an exponential RV X *with* $E[X] = 10,000$ *miles. If we want to take a 5000-mile trip, what is the probability that we can complete it without having to replace any tires?*

INTERESTING FEATURE: We don't need to assume that the tires are new when we start, because of memorylessness.

(Does this e.g. fit the general situation we talked about? Partly yes, partly no...)

Other models for waiting times

Suppose we use a RV to model the lifetime of some device.

If we use an exponential RV, then the memoryless property says that, if the device is still working at time t, then the probability that it survives another s units of time is the same as for a brand new device. That is, the device has no 'memory' of how much time as passed.

When might this be a reasonable model of a device lifetime?

- If failure is by gradual wear (like a car battery): NO
- If failure is by a catastrophe from outside (like dropping your cellphone): maybe YES.

(This doesn't stop Ross and others from making up unrealistic e.g.s: consider his original E.g. 5.5b or 5.5d.)

A good way to describe more a general waiting-time model is by the following calculation.

Suppose X is any non-negative continuous RV that we are using to model the lifetime of some item, say in seconds.

Let F and f be the CDF and PDF of X.

Then for any $t \geqslant 0$, we ask

> *What is the probability of failure in the next tiny interval of dt seconds, given that it has already survived t seconds?*

In notation, this is the quantity

$$P(t < X \leqslant t + dt \mid X > t) = \frac{P(t < X \leqslant t + dt)}{P(X > t)} \approx \frac{f(t)\,dt}{1 - F(t)}.$$

The function

$$\underbrace{\lambda(t)}_{\text{notation}} = \frac{f(t)}{1 - F(t)}$$

measures this conditional probability relative to the length of the interval, dt.

It is called the **hazard rate function** of X. It is defined as long as $F(t) < 1$, i.e. as long as failure by time t is not yet certain.

Example

If X is $Exp(\lambda)$, then

$$\boxed{\lambda(t) = \frac{\lambda e^{-\lambda t}}{1-(1-e^{-\lambda t})} = \lambda \quad \textit{for all } t \geqslant 0.}$$

This is actually another way of expressing the memoryless property: the probability of failure in the next dt, given survival until time t, doesn't depend on t.

What about other situations? Some intuitive guidance:

- If you're modeling something that wears out over time, the hazard rate function should <u>increase</u> with t: an old item should be more likely to fail soon than a new one.
- Sometimes, very new items are actually more vulnerable, and then the hazard rate function should <u>decrease</u> with t. Example: models of infant mortality.

In practice, many models are chosen by first choosing the hazard rate function to fit the data, and then reconstructing the CDF/PDF from the hazard rate function. See, e.g., Ross pp202–203.

Example (The Gamma dist$^{\text{n}}$; Ross Subsection 5.6.1)

*A RV X has a **Gamma distribution** with parameters $\alpha > 0$ and $\lambda > 0$ if $X \geqslant 0$, X is continuous, and its PDF is*

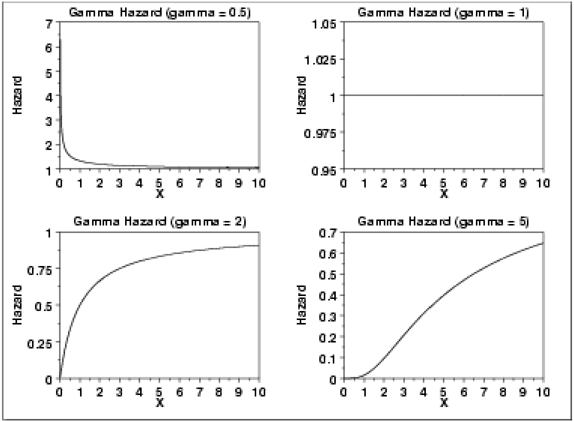
$$f(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \quad \text{for } x \geqslant 0.$$

The normalizing constant here is the so-called 'Gamma function':

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1}\, dy.$$

Observe: Gamma(α, λ) is just Exp(λ) when $\alpha = 1$. Other values of α give a generalization.

The hazard rate function of a Gamma(α, λ) RV is constant for $\alpha = 1$ (since then just Exp(λ)), increasing for $\alpha > 1$ and decreasing for $\alpha < 1$.



See Ross for more on when one uses Gamma RVs.