SV – Probability and Statistics I

http://moodle.epfl.ch/course/view.php?id=14074

Lecture 5

- Continuous RVs
- Uniform distribution
- Normal (Gaussian) distribution
- *Normal approximation to the binomial distribution*

Continuous RVs

- A random variable that can take on at most a *countable* number of possible values is a **discrete** RV
- We can also consider a random variable X whose set of possible values is *uncountable*
- X is a **continuous** RV if there exists a nonnegative function f defined on (∞, ∞) such that for any (measurable) set B of real numbers

$$P(X \in B) = \int_B f(x)dx$$

- The function f is the **probability density function (pdf)** of the RV X
- The pdf is analogous to the pmf for a discrete RV

Probability = Area under the curve

Recall some definitions :

- The **sample space** of the experiment is the set of all possible results
- Any subset of the sample space is called an **event**
- For a continuous RV X, the event that the value of X is in the set B (a subset of the sample space) is written mathematically as : $X \in B$
- The *probability* that the value x of $X \in B$ (*i.e.*, the probability of the event B) is obtained from the integral of the density f = *area under the curve*

Properties of a pdf

- The RV X *must* take on some value in $(-\infty, \infty)$, so the pdf f must satisfy

$$P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x)dx = 1.$$

- Just as for a discrete RV with pmf $p(x)$
$\implies$ *we can answer any probability question for a continuous RV if we know its pdf f*

Illustration using the pdf

- Many questions are of the form :
  'What is the probability that X is between a and b (inclusive) ?, *i.e.*, the set $B = [a, b]$:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

- *The probability that the RV X is equal to a specific value a* :

$$P(X = a) = \int_a^a f(x)dx = \underline{\underline{\qquad}}$$

$$P(X = a) = \int_a^a f(x)dx = \underline{\underline{\quad 0 \quad}}$$

- That is :

only intervals can have positive probability

Cumulative distribution function (cdf)

- Just as for discrete RVs, the **cumulative distribution function (cdf)** of a continuous RV as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

- $P(X < x) = \ldots$

Relationship between pdf and cdf

- The relationship between the cdf F and pdf f is given by

$$F(a) = P(X \in (-\infty, a]) = \int_{-\infty}^{a} f(x)dx$$

- *Differentiating* (taking the derivative) of both sides of the above equation gives

$$\frac{d}{da}F(a) = f(a)$$

- That is, the density function is *the derivative* of the cumulative distribution function

Example

Example 5.2 : The lifetime in hours of a kind of radio tube is a random variable with density

$$f(x) = \begin{cases} 0 & x \leq 100 \\ \frac{100}{x^2} & x > 100. \end{cases}$$

Assuming that tubes need replacement independently, what is the probability that exactly 2 of 5 tubes in a radio will have to be replaced within the first 150 hours of operation ??

Solution

1. Let X = number of tubes that need replacement within the first 150 hours of operation

2. $X \sim \; Bin(n = 5, \, p = \; ??)$; **Verify the 4 conditions** :

[i] *fixed* $n \; (= 5)$

[ii] *Bernoulli trials (replacement < 150/not)*

[iii] *independent*

[iv] *same probability of defective* $p \; (= \; ??)$

Solution, cont.

$$\mathbf{p} = \int_0^{100} 0\,dx + \int_{100}^{150} \frac{100}{x^2}\,dx = 0 - \frac{100}{x}\Bigg|_{100}^{150}$$

$$= -\frac{100}{150} - \left(-\frac{100}{100}\right) = 1 - \frac{2}{3} = \mathbf{\frac{1}{3}}$$

3. $P(X = 2)$

4. $= \binom{5}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = \frac{80}{243}$ $\boxed{\approx \underline{\underline{0.329}}}$

[subst. binom. $p(i)$, simp.]

Expected value of a continuous RV

- For a *continuous* RV X, the **expected value** is just the continuous analogue of the sum :

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

- For a real-valued function g, $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$
- It is also straightforward to show for constants a and b,

$$E[aX + b] = aE[X] + b$$

Variance of a continuous RV

- The **variance** for a continuous RV is defined in exactly the same way as for a discrete RV :

$$Var(X) = E[(X - E[X])^2]$$

- Again, we can also use the alternative formula :

$$Var(X) = E[X^2] - (E[X])^2$$

- For constants a and b, we also have

$$Var[aX + b] = a^2 Var(X)$$

Uniform distribution

- A RV X is **uniformly distributed** on the interval (α, β) if its density is

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & \text{otherwise.} \end{cases}$$

Example 5.3 Find the cdf F for $X \sim U(\alpha, \beta)$ ...

Expected value for a uniform RV

- If $X \sim U(\alpha, \beta)$:

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} \\
&= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\
&= \boxed{\frac{\beta + \alpha}{2}}
\end{aligned}
$$

- $\Rightarrow$ The expected value of a RV uniformly distributed on an interval is equal to the midpoint of the interval

Variance of a uniform RV

We will use the alternative formula ; first we calculate

$$E[X^2] = \int_\alpha^\beta \frac{x^2}{\beta - \alpha}$$

$$= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)}$$

$$= \frac{\beta^2 + \alpha\beta + \alpha^2}{3}$$

Thus

$$Var(X) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \frac{(\alpha + \beta)^2}{4} = \boxed{\frac{(\beta - \alpha)^2}{12}}$$

Example

Example 5.3 $X \sim U(0,1) \Rightarrow f(x) =$ ____________________

(a) $P(X < .3) =$

(b) $P(X > .6) =$

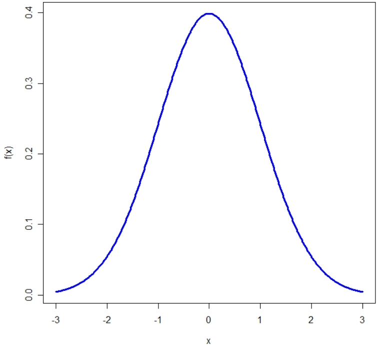
(c) $P(.3 < X < .8) =$

PAUSE

Normal distribution

- A RV X is **Normally distributed** (or has a **Gaussian** distribution) with parameters μ and σ^2 if the density of X is given by

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



Integral of the Normal density = 1

- (optional : this will not be examined ! !)
- In order to prove that $f(x)$ is a true probability density, we must show that

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

- By making a change of variables $y = (x - \mu)/\sigma$, we obtain

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2} dy$$

- Thus, we must show that $\int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi}$

Integral = 1, cont.

- Soit $I = \int_{-\infty}^{\infty} e^{-y^2/2} dy$:

$$I^2 = \int_{-\infty}^{\infty} e^{-y^2/2} dy \int_{-\infty}^{\infty} e^{-x^2/2} dx$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y^2+x^2)/2} dy\, dx,$$

- By using polar coordinates, we can evaluate this double integral :

$$x = r\, cos\theta,\ \ y = r\, sin\theta,\ \ dy\, dx = r\, d\theta\, dr$$

- Then

$$I^2 = \int_0^{\infty} \int_0^{2\pi} e^{-r^2/2} r\, d\theta\, dr$$
$$= 2\pi \int_0^{\infty} r e^{-r^2/2} dr$$
$$= -2\pi e^{-r^2/2} \Big|_0^{\infty} = 2\pi \quad => \quad I = \sqrt{2\pi}$$

Expected value of a Normal RV

- To find $E[X]$, $X \sim N(\mu, \sigma^2)$, we must calculate

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

- We won't work out the details here, but re-writing the multiplier x as $(x - \mu) + \mu$ we can split the integral into 2 parts
  - one part is μ times the integral of a normal density (which is therefore equal to 1)
  - the other part is symmetric around 0 so the positive and negative pieces cancel and the integral is 0
- So for $X \sim N(\mu, \sigma^2)$, $E[X] = \mu$

Variance of a Normal RV

- To find $Var[X]$, $X \sim N(\mu, \sigma^2)$, we use the definition :

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx. \end{aligned}$$

- This can be done very simply, using the substitution $y = (x-\mu)/\sigma$ then using integration by parts
- Thus, if $X \sim N(\mu, \sigma^2)$, $Var(X) = \sigma^2$

Distribution of Y = aX + b

- An important and useful fact about Normal RVs (that is *NOT TRUE* for all RVs) :

  $$\textit{if } X \sim N(\mu, \sigma^2), \textit{ then } Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$$

- To prove this result, we can find the cumulative distribution function F of the RV $Y = aX + b$; $X \sim N(\mu, \sigma^2)$, and a, b constants

- Differentiating F gives the density of Y, which is of the form of a normal density

Normal RV in standard units

- The most useful application of the preceding result is in determining probabilities for Normally distributed RVs
- If $X \sim N(\mu, \sigma^2)$, the RV $Z = (X - \mu)/\sigma \sim N(0, 1)$
- The distribution of Z is the **standard Normal distribution**
- We denote the density function $f(z)$ by the symbol ϕ, and the cumulative distribution function Φ :

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

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz$$

- Since this integral does not have a simple form, we use a *table of computed values* (or computer software) for finding the *area under the curve*

Normal table

TABLE 5.1 AREA $\Phi(x)$ UNDER THE STANDARD NORMAL CURVE TO THE LEFT OF x

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Solving problems using the Normal table

- Steps to solving problems involving *Normally distributed* RVs (*i.e.*, Step 4 : 'Answer the question' for $X \sim N(\mu, \sigma^2)$) :

1. Convert the RV to *standard units (SU)*
2. **** DRAW THE PICTURE ****
3. Use the *Normal Table* to find the necessary probabilities
4. (Do the calculations)

Example : practice using the Normal table

Example 5.4 Let X be a Normal RV with parameters $\mu = 3$ and $\sigma^2 = 9$. Calculate :

(a) $P(2 < X < 5)$

Example, part (b)

(b) $P(X > 0)$

Another example

Example 5.5 Let $X \sim N(\mu = 66, \sigma^2 = 9^2)$. Find the number c such that $P(X < c) = 0.75$: