MAT2377

Catalin Rada

Version May 14, 2009

Lectures

Ch. 4 Continuous Random Variables

What to do when there are uncountably infinitely many outcomes, e.g. points on the real line?

In the discrete case, the probability mass function $P(X = x)$ was the main object of interest. In the continuous case this role will be played by the **probability density function**.

The **(cumulative) distribution function** (CDF) of any rv X is

$$F_X(x) = P(X \leq x)\,,$$

viewed as a function of a real variable x.

Area under a curve

We can visualise the distribution in terms of a **probability density function** (PDF) $f_X(x) = \frac{d}{dx} F_X(x)$; then since for any $a < b$,

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

we have

$$\begin{aligned} P(X \leq a) + P(a < X \leq b) &= P(X \leq b) \\ P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx \end{aligned}$$

The last expression is just area under $f_X(x)$, from a (open) to b (closed), i.e., $(a, b]$. Because of this interpretation in terms of area, we clearly GET: $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$.

Properties of p.d.f. f_X of a continuous rv X are:

1) $f_X(x) \geq 0$ for all x;

2) $\int_{-\infty}^{\infty} f_X(x)dx = 1$;

3) $P(a \leq X \leq b) = \int_a^b f_X(x)dx$ - the area under f_X from a to b.

Example 10:

(A) Assume that X has the following p.d.f.

$$f_X(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \leq x \leq 2 \\ 0 & x > 2 \end{cases}$$

The corresponding c.d.f. is given by:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \int_0^x f_X(z)\,dz = \frac{1}{2}\int_0^x z\,dz = \frac{x^2}{4} & 0 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

$$P(0.5 < X < 1.5) = F_X(1.5) - F_X(0.5) = \frac{(1.5)^2}{4} - \frac{(0.5)^2}{4}$$

(B) Assume that X has the following p.d.f.

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & x < 0 \end{cases}$$

The corresponding c.d.f. is given by:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \int_0^x f_X(z)\, dz = \lambda \int_0^x \exp(-\lambda z)\, dz = 1 - \exp(-\lambda x) & x > 0 \end{cases}$$

(C) Do at the blackboard exercise 4-4/page 113!

(D) Do at the blackboard exercise 4-12/page 116!

Ok, let me type the solutions of these Calculus exercises....

(C) a) $P(1 < X) = \int_1^\infty f_X(x)dx = \int_4^\infty f_X(x)dx$ since $f_X(x) = 0$ for $x < 4$. Hence $P(1 < X) = \lim_{t\to\infty} \frac{e^{-(x-4)}}{-1}|_4^t = \lim_{t\to\infty} \frac{e^{-(t-4)}}{-1} + \frac{e^0}{1} = 1$; b) $P(2 \le X < 5) = P(2 \le X \le 5) = \int_2^5 f_X(x)dx = \int_4^5 f_X(x)dx = \int_4^5 e^{-(x-4)}dx = \frac{e^{-(x-4)}}{-1}|_4^5 = -e^{-1} + 1 \cong 0.6321$; c) $P(5 < X) = \int_5^\infty f_X(x)dx = \int_5^\infty e^{-(x-4)}dx = \frac{e^{-(x-4)}}{-1}|_5^\infty = e^{-1} \cong 0.3679$ or you may use easily b); d) $P(8 < X < 12) = \int_8^{12} f_X(x)dx = \frac{e^{-(x-4)}}{-1}|_8^{12} = -e^{-8} + e^{-4} = etc$; e) We just need solve the equation $P(X < x) = 0.90$, so we have: $\int_4^x f_X(z)dz = 0.90$, so $\int_4^x e^{-(z-4)}dz = 0.90$, hence $\frac{e^{-(z-4)}}{-1}|_4^x = 0.90 \Rightarrow \frac{e^{-(x-4)}}{-1} + 1 = 0.90$, hence $0.10 = e^{-(x-4)}$, so $\ln 0.10 = -(x-4)$, thus $x = 4 - \ln 0.10 = etc.$

(D) a) Since we do have a continuous rv we get that $P(X < 1.8) = P(X \le 1.8) = F_X(1.8) = (0.25 \times 1.8) + 0.5$, since $1.8 \in [-2, 2)$. So $P(X < 1.8) = etc...$; b) $P(X > -1.5) = 1 - P(X \le -1.5) = 1 -$

$F_X(-1.5) = 1 - \{0.25 \times (-1.5) + 0.5\} = etc$, since $-1.5 \in [-2, 2)$; c) $P(X < -2) = P(X \leq -2) = 0.25 \times (-2) + 0.5 = 0$ (just look at the definition of F_X); d) $P(-1 < X < 1) = F_X(1) - F_X(-1) = \{0.25 \times 1 + 0.5\} - \{0.25 \times (-1) + 0.5\} = etc$

Mean and variance

For a continuous rv X with pdf $f_X(x)$, the expectation is ***defined*** as

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx\,.$$

Example: Find the expected value in the example (A) above.

Solution: $\mathrm{E}(X) = \int_0^2 x f_X(x)\, dx = \int_0^2 \frac{1}{2} x^2\, dx = \frac{4}{3}$.

Expectation of a function of a cts rv; Variance

In a similar way to the discrete case,

$$\mathrm{E}\left[h(X)\right] = \int_{-\infty}^{\infty} h(x) f_X(x)\, dx\,.$$

In particular, as before,

$$\mathrm{Var}(X) \stackrel{\text{def}}{=} \mathrm{E}\left(\left[X - \mathrm{E}(X)\right]^2\right) \stackrel{\text{comp. formula}}{=} \mathrm{E}\left(X^2\right) - \left[\mathrm{E}(X)\right]^2\,.$$

Also as before we define $\mathit{SD}(X) = \sqrt{\mathrm{Var}(X)}$.

Linear function $a + bX$

If we take a linear function of a cts rv, as in the discrete case we get

$$
\begin{aligned}
\mathrm{E}\,(a + bX) &= a + b\mathrm{E}(X)\,, \\
\mathrm{Var}\,(a + bX) &= b^2\mathrm{Var}(X)\,, \\
\mathit{SD}\,(a + bX) &= |b|\mathit{SD}(X)\,.
\end{aligned}
$$

Do at the blackboard: 4-26/page 118; 4-30.

4-6 Normal Distribution

An important example is afforded by the special PDF

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ for } \quad \text{all } x \in \mathbf{R}.$$

The corresponding CDF is denoted by

$$\Phi(x) = \int_{-\infty}^{x} \phi(z)\, dz\,.$$

A rv with this CDF is said to have a **standard normal distribution**. Such a rv is traditionally denoted (where sensible) by Z, and we write $Z \sim \mathcal{N}(0, 1)$. What is the graph?

Standard Normal Random Variable

The expectation and variance of $Z \sim \mathcal{N}(0,1)$ are

$$\mathrm{E}(Z) = \int_{-\infty}^{\infty} z \, \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, dz = 0,$$

$$\int_{-\infty}^{\infty} z^2 \phi(z) \, dz = 1$$

Hence if $Z \sim \mathcal{N}(0,1)$ then

$$\mathrm{E}(Z) = 0 \;\; \text{and } \mathrm{Var}(Z) = 1 = \mathit{SD}(Z)\,.$$

The general normal RV

Take a positive σ and any real μ. Then if $Z \sim \mathcal{N}(0,1)$ and $X = \mu + \sigma Z$ then

$$\frac{X-\mu}{\sigma} = Z \sim \mathcal{N}(0,1)$$

and the CDF of X (since $\sigma > 0$) is given by

$$F_X(x) = P\{X \leq x\} = P\{\mu + \sigma Z \leq x\} = P\left\{Z \leq \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

The PDF is then

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$$

A rv X with this CDF/PDF has

$$\mathrm{E}(X) = \mu + \sigma\mathrm{E}(Z) = \mu \quad , \quad \mathrm{Var}(X) = \sigma^2\mathrm{Var}(Z) = \sigma^2 \ \Rightarrow \ \mathit{SD}(X) = \sigma$$

and is said to be **Normal with Mean (expectation) $\boldsymbol{\mu}$ and Variance $\boldsymbol{\sigma^2}$**; write $X \sim \mathcal{N}(\mu, \sigma^2)$.

Examples

1. Assume that Z represents the standard normal random variable. Evaluate the following probabilities:

   (a) $P(Z \leq 0.5) = 0.6915$
   (b) $P(Z < -0.3) = 0.3821$
   (c) $P(Z > 0.5) = 1 - P(Z \leq 0.5) = 1 - 0.6915 = 0.3085$,
   (d) $P(0.1 < Z < 0.3) = P(Z < 0.3) - P(Z < 0.1) = 0.6179 - 0.5398 = 0.0781$,
   (e) $P(-1.2 < Z < 0.3) = P(Z < 0.3) - P(Z < -1.2) = 0.5028$.

2. Suppose that the waiting time (in minutes) for a coffee at 9am is normally distributed with mean 5 and standard deviation 0.5. What is the probability that one such waiting time is at most 6 minutes?

Solution: Let X denote the waiting time; then $X \sim \mathcal{N}(5, 0.5^2)$ and the ***standardised rv***

$$Z = \frac{X-5}{0.5} \sim \mathcal{N}(0,1)\,.$$

The desired probability is

$$P\{X \leq 6\} = P\left\{\frac{X-5}{0.5} \leq \frac{6-5}{0.5}\right\} = P\left(Z \leq \frac{6-5}{0.5}\right) = \Phi\left(\frac{6-5}{0.5}\right) = \Phi(2)$$

Reading from the tables, $\Phi(2) \approx 0.9772\,.$

3. Suppose that bottles of beer are filled in such a way that the actual volume in them (in mL) varies randomly according to a normal distribution with mean 376.1 with standard deviation 0.4. What is the probability that the volume is less than 375mL?

   Solution: Let X denote the volume; then $X \sim \mathcal{N}(376.1, 0.4^2)$ and so

$$Z = \frac{X - 376.1}{0.4} \sim \mathcal{N}(0,1)\,.$$

   The desired probability is

$$P\left(X < 375\right) = P\left(\frac{X - 376.1}{0.4} < \frac{375 - 376.1}{0.4}\right) = P\left(Z < \frac{-1.1}{0.4}\right)\,.$$

   So

$$\Phi\left(-2.75\right) = 0.003\,.$$

Reading table backwards

In some examples we are asked to go in the other direction: If $Z \sim \mathcal{N}(0,1)$, for which values a, b and c do we have

- $P(Z \leq a) = 0.95$;
- $P(|Z| \leq b) = P(-b \leq Z \leq b) = 0.99$;
- $P(|Z| \geq c) = 0.01$.

Solution:

- Observing the table we see that

$$P(Z \leq 1.64) \approx 0.9495 \quad \text{and} \quad P(Z \leq 1.65) \approx 0.9505\,.$$

Clearly we must have $1.64 < a < 1.65$; linearly interpolating, our best guess would be $a \approx 1.645$, although this level of precision is usually not necessary. Often the inital interval estimate is enough.

- Next, note that

$$P\,(-b \leq Z \leq b) = P(Z \leq b) - P(Z < -b)$$

However by symmetry etc. we get that

$$P(Z < -b) = P(Z > +b) = 1 - P(Z \leq +b)$$

and so

$$P\left(-b \leq Z \leq b\right) = P(Z \leq b) - [1 - P(Z \leq b)] = 2P(Z \leq b) - 1$$

If this is 0.99, then

$$P(Z \leq b) = \frac{1 + 0.99}{2} = 0.995\,;$$

Consulting the table we see that

$$P(Z \leq 2.57) \approx 0.9949 \quad \text{and} \quad P(Z \leq 2.58) \approx 0.9951$$

so again an interpolation suggests taking $b \approx 2.575$.

- Finally note that $\{|Z| \geq c\} = \{|Z| < c\}^c$ so we need c such that

$$P\left(|Z| < c\right) = 1 - P\left(|Z| \geq c\right) = 0.99$$

But this is the same as

$$P\left(-c < Z < c\right) = P(-c \leq Z \leq c) = 0.99$$

since $|x| < y \Leftrightarrow -y < x < y$, and $P(Z = c) = 0$ for all c. Hence again we take $c \approx 2.575$.

Some exercises:

1) The length for chip manufacturing is assumed to follow the normal distribution with a mean of 0.5 mm and a standard deviation of 0.05 mm. i) What is the probabbility that a length is greater than 0.62 mm? ii)

What is the chance that a length is between 0.47 mm and 0.63 mm? iii) The length of 90% of samples is below what value?

Sol: i) $P(L > 0.62) = 1 - P(L \leq 0.62) = 1 - P(\frac{L-0.5}{0.05} \leq \frac{0.62-0.5}{0.05}) = 1 - P(X \leq 2.4) = 1 - \Phi(2.4) = 1 - 0.991802 = 0.008198 \cong 0.0082$; here $X \sim \mathcal{N}(0,1)$. ii) $P(0.47 < L < 0.63) = P(\frac{0.47-0.5}{0.05} < \frac{L-0.5}{0.05} < \frac{0.63-0.5}{0.05}) = P(-0.6 < X < 2.6) = \Phi(2.6) - \Phi(-0.6) = 0.995339 - 0.274253 = 0.721086 \cong 0.72109$; here $X \sim \mathcal{N}(0,1)$. iii) We need to find x such that $P(L \leq x) = 0.90$. We note that: $0.90 = P(L \leq x) = P(\frac{L-0.5}{0.05} \leq \frac{x-0.5}{0.05})$; so $\frac{x-0.5}{0.05}$ should be the middle of (see tables) the interval $[1.28, 1.29]$, i.e., 1.285. Hence $x = (0.05 \times 1.285) + 0.5 = 0.56425 \cong 0.5642$.

2) The height in inches of a population of aliens from planet MathematiX is normally distributed with mean 69.5 and standard deviation 2.5. What is the probability that an alien chosen randomly from planet MathematiX is i) over 72 inches tall? ii) between 69.5 and 73.1 inches tall; iii) less than 67 inches tall?

Sol: i) Let H be the height! It is a continuous rv. We need to compute $P(H > 72)$, so we do it as follows: $P(H > 72) = 1 - P(H \leq 72) = 1 - P(\frac{H-69.5}{2.5} \leq \frac{72-69.5}{2.5}) = 1 - P(X \leq 1) = 1 - \Phi(1) = 1 - 0.841345 = 0.158655$ from the table on page 713. ii) We need to compute $P(69.5 < H < 73.1)$, so since it is a continuous rv we get $P(69.5 < H \leq 73.1) = P(\frac{69.5-69.5}{2.5} < \frac{H-69.5}{2.5} \leq \frac{73.1-69.5}{2.5}) = P(0 < X \leq 1.44) = \Phi(1.44) - \Phi(0) = 0.925066 - 0.500000 = 0.425066$, where X is having a standard normal distribution; iii) We need to compute $P(H < 67) = P(H \leq 67) = P(\frac{H-69.5}{2.5} \leq \frac{67-69.5}{2.5}) = P(X \leq -1) = \Phi(-1) = 0.158655$, where X is having a standard normal distribution!

4-8 Exponential random variable

Assume that cars arrive according to a Poisson process, i.e. number of cars, say N, in 1 hour is a Poisson rv with parameter λ. Let X be the time to the first car arrival, it is a rv. What is $P(X > x)$? What do we have? i) for 1 hour the rate is λ; ii) for x hours the rate is $\frac{\lambda \times x}{1} = \lambda x$. Then

$$P(X > x) = P(N = 0) = \exp(-\lambda x) \times \frac{(\lambda x)^0}{0!} = e^{-\lambda x}.$$

We say that X has the exponential distribution with parameter λ, $X \sim$ exponential, λ.

- $$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\lambda x} & \text{for } 0 \leq x \end{cases}$$

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

- $\mu = \mathrm{E}(X) = 1/\lambda$

- $\sigma^2 = \mathrm{Var}(X) = 1/\lambda^2$

- "Lack of memory":

$$P(X > s + t \mid X > t) = P(X > s)$$

Example: The lifetime of a certain type of lightbulb has an exponential distribution with mean 100 hours.

1. What is the probability that a lightbulb will last at least 100 hours?

2. Given that a lightbulb has already been burning for 100 hours, what is the probability that it will last at least 100 hours more?

3. The manufacturer wants to guarantee that his lightbulbs will last at least t hours. What should t be in order to ensure that 90% of the lightbulbs will last longer than t hours?

Answer: X is exponential with $\lambda = 1/100$; indeed $100 = \frac{1}{\lambda}$ To compute 1) $P(X > 100) = e^{-(1/100)100} = e^{-1} = etc$; 2) $P(X > 200|X > 100) = P(X > 100)$; 3) to find t such that $P(X > t) = 0.9$, so $e^{-\lambda \times t} = 0.9$, so $-\lambda t = \ln(0.9)$, so $t = 10.5360$.

4.9 Gamma random variable

Assume that cars arrive according to a Poisson process with rate λ. Recall that if X is the time to the first car arrival, then X has exponential distribution with parameter λ. Now, if X is the time to the rth arrival, then X has **Gamma distribution** with parameters λ and r. We have

- $$\mu = \mathrm{E}(X) = \frac{r}{\lambda}$$

- $$\sigma^2 = \mathrm{Var}(X) = \frac{r}{\lambda^2}$$

So for $r = 1$ one gets back the exponential distribution! The p.d.f. in this case is $f_X(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}$ where $r = 0, 1, 2, 3, \ldots$ and $x > 0$. The c.d.f. is given by $F_X(x) = P(X \leq x) = 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!}$

EXC: Some rocks from the planet MathematiX are analyzed for contamination. We assume that the number of particles of contamination per kg of rocks is a Poisson rv with mean 0.01 particle per kg. i) Compute the expected number of kgs of rocks required to get 15 particles of contamination? ii) What is the standard deviation of the kgs of rocks required to get 15 particles of contamination?

Sol: i) Let R be the kgs of rocks to get 15 particles. From the statement we get that R is a rv with Gamma distribution: $r = 15$, and $\lambda = 0.01$. By the formula (on the previous page) we get $E(R) = \frac{r}{\lambda} = \frac{15}{0.01} = 1500$ kgs; ii) We get first the variance, and after that the stand. deviation, as follows (see the same page): $Var(R) = \frac{r}{\lambda^2} = \frac{15}{(0.01)^2} = 150000$, so the standard deviation is $\sqrt{150000} = etc.$

Exc. 1 Buses arrive in a certain bus station according to a Poisson process at a rate of 20 per hour. What is the probability that a passenger will have to wait more than 5 minutes for the arrival of the first bus?

Sol: PLAN: we need a rv, then indentify the distribution, then finish the computations! Let Y be the waiting time (in minutes!) until the first bus appears in the bus station. We need to compute $P(Y > 5)$. The distribution is exponential! So we need to find our new rate: if for 1 hour = 60 minutes we have a rate of 20 buses, for 1 minute we have a rate of $\frac{20\times1}{60} = \frac{1}{3}$; so $P(Y > 5) = \int_5^\infty f_Y(x)dx = \int_5^\infty \frac{1}{3}e^{-\frac{1}{3}x}dx = \frac{1}{3} \times \frac{e^{-1/3}}{-1/3}|_5^\infty = -[0 - e^{-5/3}] = e^{-5/3} = etc.$

Exc. 2 Airplanes arrive in a certain airport according to a Poisson process at a rate of 30 airplanes per hour. What is the chance that a traffic controller will wait more than 5 minutes before both of the first 2 airplanes arrive?

Sol: Same plan! Let Y be the rv denoting the waiting time (in minutes) until the second airplane arrives. Then Y has Gamma distribution with $r = 2$ and rate $=$? If for 1 hour $=$ 60 minutes the rate is 30 airplanes, then the for 1 minute the rate is $\frac{1}{2}$. We need $P(Y > 5) = \sum_{k=0}^{2-1} \frac{e^{\frac{-5}{2}} \times (\frac{5}{2})^k}{k!} = \frac{e^{\frac{-5}{2}} \times 1}{1} + \frac{e^{\frac{-5}{2}} \times \frac{5}{2}}{1} = e^{\frac{-5}{2}}[1 + \frac{5}{2}] = e^{\frac{-5}{2}}[7/2] \cong 0.28.$

Normal approximation to the binomial

If $X \sim \mathcal{B}(n,p)$ then *we may interpret* X as a sum of iid rv's:

$$X = I_1 + I_2 + \cdots + I_n \text{ where each } I_i \sim \mathcal{B}(1,p)\,.$$

Thus according to the Central Limit Theorem, for large n, the standardised version

$$\frac{X - np}{\sqrt{np(1-p)}} \overset{\text{approx}}{\sim} \mathcal{N}(0,1)\,,$$

i.e. for large n if $X \overset{\text{exact}}{\sim} \mathcal{B}(n,p)$ then(pg13-14) $X \overset{\text{approx}}{\sim} \mathcal{N}(np, np(1-p))$.

This means that binomial probabilities can be approximated by normal probabilities. In other words: $P(X \leq x) \cong \Phi(\frac{x-np}{\sqrt{np(1-p)}})$.

Discrete vs. Continuous

4.7 Normal approximation to binomial with continuity correction

Let $X \sim \mathcal{B}(n, p)$. Recall that $\mathrm{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$. If $np > 5$ and $n(1-p) > 5$, we may approximate X by a normal random variable in the following way:

$$P(X \leq x) = P(X < x + 0.5) \cong P\left(Z < \frac{x - np + 0.5}{\sqrt{np(1-p)}}\right)$$

and

$$P(X \geq x) = P(X > x - 0.5) \cong P\left(Z > \frac{x - np - 0.5}{\sqrt{np(1-p)}}\right).$$

Examples:

1. Suppose $X \sim \mathcal{B}(36, 0.5)$. Provide a normal approximation to the probability $P(X \leq 12)$. *Note:* For $n = 36$ the binomial probabilities are not available in the textbook.

   Sol: Since $\mathrm{E}(X) = 36 \times 0.5 = 18 > 5$, $\mathrm{Var}(X) = 36 \times 0.5 \times 0.5 = 9 > 5$,

$$P(X \leq 12) = P\left(\frac{X - 18}{3} \leq \frac{12 - 18 + 0.5}{3}\right) \overset{\text{norm.approx'n}}{\approx} \Phi(-1.83) \overset{\text{table}}{\approx} 0.033\,.$$

   Compare this to the exact value (from computer) of 0.0326.

Computing binomial probabilities - summary

We thus have at least 3 ways to compute (or approximate) binomial probabilities:

- Use the exact formula: if $X \sim \mathcal{B}(n,p)$ then for each $x = 0, 1, \ldots, n$, $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$;
- Use tables: if $n \leq 15$ and p is one of 0.1, 0.2,…, 0.9 then the CDF is in the textbook (must express desired probability in terms of CDF, i.e. in form $P(X \leq x)$ first), i.e.

$$P(X < 3) = P(X \leq 2) \quad ; \quad P(X = 7) = P(X \leq 7) - P(X \leq 6)\,;$$

$$P(X > 7) = 1 - P(X \leq 7) \quad ; \quad P(X \geq 5) = 1 - P(X \leq 4) \text{ etc.}$$

- Use normal approximation. The “rule of thumb” in the binomial case becomes: if np and $n(1-p)$ are both ≥ 5, the normal approximation, i.e. for $x = 0, 1, \ldots, n$, if $X \sim \mathcal{N}(np, np(1-p))$,

$$P(X \leq x) \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(X \geq x) \approx 1 - \Phi\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

should provide an accurate approximation.

EXC 1. Suppose that the probability a soccer player has an accident during a one year period is 0.4. If there are 200 soccer players in a city, compute the chance that more than 75 players will have an accident in one year. We assume: that a player is getting accidented or not is independent of any other player getting accidented!

Sol: Set X be the rv: the number of players that are getting an accident during a year. Then X is a binomial rv with $n = 200$, $p = 0.4$, $1 - p = 0.6$. So $E(X) = np = 200 \times 0.4 = 80 > 5$, and standard dev. $\sqrt{200 \times p \times (1-p)} = \sqrt{200 \times 0.4 \times 0.6} = \sqrt{48} = 6.928$. We have to calculate $P(X > 75)$, but we are going to approximate (since computtaions are too long...). So (by page 33) $P(X > 75) = P(X \geq 76) = P(X > 76 - 0.5) \cong P(\frac{X-np}{\sqrt{np(1-p)}} > \frac{76-0.5-np}{\sqrt{np(1-p)}}) = P(Z > \frac{-4.5}{6.928}) = P(Z > -0.65) = 1 - P(Z \leq -0.65) = 1 - 0.257846 = 0.742154 \cong 0.7422$.

EXC 2. A new type of coffee is tested on 100 students (during exam period) to determine its effects. It is claimed that a random student who drinks this new type of coffee will get A- with probability 0.8; i) Compute the probability that less than 74 students are getting A-; ii) Compute the chance that between 74 and 85 students, inclusive, are going to get A-.

Sol: Set X be the number of students that are getting A-. It is a binomial rv with $n = 100$, $p = 0.8$, $1 - p = 0.2$, $E(X) = np = 100 \times 0.8 = 80 > 5$, standard dev. being $\sqrt{np(1-p)} = \sqrt{80 \times 0.2} = 4$. Note that $n(1-p) > 5$. i) By page 33 (again) we get: $P(X < 74) = P(X \leq 73) = P(X < 73 + 0.5) \cong P(Z < \frac{73+0.5-np}{\sqrt{np(1-p)}}) = P(Z < \frac{73.5-80}{4}) = P(Z < -1.62) = P(Z \leq -1.62) = 0.052616$ (from the table on page 712); ii) We need $P(74 \leq X \leq 85)$. By the same page 33 (2 times now!) we get $P(74 \leq X \leq 85) \cong P(\frac{73.5-80}{4} < Z < \frac{85.5-80}{4}) = P(-1.63 < Z < 1.37) = \Phi(1.37) - \Phi(-1.63) = 0.914657 - 0.051551 = 0.863106 \cong 0.8631$.

Section 4.5 **Continuous uniform distribution**

Definition: A continuous uniform rv is a continuous rv X whose p.d.f. is given by $f_X(x) = \frac{1}{b-a}$ for $a \leq x \leq b$.What is the graph? What is the expectation? $E(X) = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\{\frac{x^2}{2}|_a^b\} = \frac{b+a}{2}$, i.e., the middle of the interval $[a, b]$. Similarly $Var(X) = \frac{(b-a)^2}{12}$.

What is the C.D.F.? $F_X(x) = P(X \leq x) = \int_a^x \frac{1}{b-a}dz = \frac{x-a}{b-a}$, for $a \leq x < b$; $F_X(x) = 0$ if $x < a$, and $F_X(x) = 1$ if $x \geq b$.

EXC: Buses on a certain street (on planet MathematiX) run every 10 minutes between rush hours. What is the probability that an alien entering the station at a random time will have to wait at least 8 minutes?

Sol: Define the rv X to be time to the next bus. This rv is a continuous $uniform$ rv on the interval $[0, 10]$ because the alien is entering the station at a random time! We need to compute $P(X \geq 8) =$

$\int_8^{10} \frac{1}{10-0} dx = 2/10 = 1/5$ — which is not too bad!!!

5-5 Independent Random Variables

Def 1. Let X and Y be 2 rv (discrete or continuous). Then X and Y are called independent if:

$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively!

Def 2. Given rv $X_1, X_2, \ldots, X_n$ and constants $a_1, a_2, \ldots, a_n$, then $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ is called a linear combination of $X_1, X_2, \ldots, X_n$.

Theorem 1 i) $E(Y) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$; ii) $Var(Y) = a_1^2Var(X_1) + a_2^2Var(X_2) + \cdots + a_n^2Var(X_n) + 2\sum\sum_{i<j} a_ia_jcov(X_i, X_j)$ where $cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$. Moreover, if $X_1, X_2, \ldots, X_n$ are

independent rv then $Var(Y) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + \cdots + a_n^2 Var(X_n)$.

As an application (plus other considerations):

Theorem 2 i) If $\bar{X} = \frac{X_1+X_2+\cdots+X_p}{p}$ and $E(X_i) = \mu$ for all $i = 1, 2, \ldots, p$, then $E(\bar{X}) = \mu$. If $X_1, X_2, \ldots, X_p$ are also independent with $Var(X_i) = \sigma^2$ for all $i = 1, 2, \ldots, p$, then $Var(\bar{X}) = \frac{\sigma^2}{p}$.

ii) Suppose that $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, 3, \ldots, p$, and suppose that $X_1, X_2, \ldots, X_p$ are independent. Then $Y = a_1 X_1 + a_2 X_2 + \cdots + a_p X_p$ is a NORMAL rv with:

mean: $E(Y) = a_1\mu_1 + a_1\mu_2 + \cdots + a_1\mu_p$ and

variance: $Var(Y) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_p^2\sigma_p^2$