Crash Course in Statistics for Neuroscience Center Zurich University of Zurich

Dr. C.J. Luchsinger

4 Selected Probability Distributions - all in 1 chapter so that you find them again

* discrete: Bernoulli, Binomial
* continuous: Uniform, Normal, χ^2, F, t
* possible values, Distribution (Probability function / Density), E, V

4.1 Discrete

4.1.1 Bernoulli Be(p)

Two possible values 0 and 1 (alternatively -1 and $+1$). $P[X=1]=p$ (Success) and $P[X=0]=1-p$ (Failure). $E[X]=p$ and $V[X]=p(1-p)$.

$$P[X=x]=p^x(1-p)^{1-x}$$

4.1.2 Binomial Bin(n,p)

X_i, $1 \le i \le n$, n iid Be(p)-R.V. $Y := \sum_{i=1}^n X_i$. Y so called Binomial RV with parameters n and p; Bin(n,p). $E[Y]=np$ and $V[Y]=np(1-p)$. Possible values: natural numbers $0 \le y \le n$:

$$P[Y=y]=\binom{n}{y}p^y(1-p)^{n-y}.$$

"How many successes in n trials?"

4.2 Continuous

4.2.1 Uniform $U[a,b]$

U is distributed uniformly on $[a,b]$, if U has following density function:

$$f(u) = (b-a)^{-1},$$

$a \leq u \leq b$. For $u \notin [a,b]$ density is 0. $E[U] = (a+b)/2$ and $V[U] = (b-a)^2/12$.

4.2.2 Normal N(μ, σ^2), also called Gauss-, Glocken-, Bell-, Forrest Gump-Distribution

Very important. Density:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

$E[X] = \mu$ and $V[X] = \sigma^2$.

When constructing Confidence Intervals (CI's) and statistical tests (Chapters 6 and 7), we often have to sum (normal) random variables:

$$X_1 + X_2 + \ldots + X_n.$$

We then need to know the expectation and variance of such sums and mostly even their distribution. The first 2 results to follow will help us find them.

DANGER: FOLLOWING RESULTS DO *NOT* HOLD FOR MANY DISTRIBUTIONS - BUT FOR THE NORMAL DISTRIBUTION IT'S OK!

1. Sum of 2 independent and normally distributed RV's is again normally distributed. Let X be $\mathcal{N}(\mu_1, \sigma_1^2)$ and Y be $\mathcal{N}(\mu_2, \sigma_2^2)$, X independent of Y. Then $X+Y$ has a

$$\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

distribution.

2. "Z-Transform": If X has a $\mathcal{N}(\mu, \sigma^2)$ distribution, then

$$\frac{X - \mu}{\sigma}$$

has a $\mathcal{N}(0,1)$-distribution (which can be found in tables in books and libraries of statistical packages). $\mathcal{N}(0,1)$ is called "standard" normal.

Compute $P[X > 13]$ when X is a $\mathcal{N}(10,4)$ using tables. See 4.3 for further exercises.

3. approximate results: for any σ: let X be $\mathcal{N}(0, \sigma^2)$ RV, then

$$P[|X| > \sigma] \doteq \frac{1}{3}$$

and

$$P[|X| > 2\sigma] \doteq 5\%.$$

Generally: Normal distribution has about 2/3 of it's probability within 1 standard deviation away from the mean (more precise, see table %) and even 95 % within 2 standard deviations away from the mean:

Due to the so called Central Limit Theorem (CLT, not treated in this course), we have an approximate result (for large n; common rule is $np(1-p) > 9$) in case of the Binomial Distribution: Almost 95 % of the probability lies in the interval

$$[np - 2\sqrt{np(1-p)} \ , \ np + 2\sqrt{np(1-p)}].$$

Following 3 distributions will be revisited when we need them in chapters 6, 7 and 8. But the χ^2-distribution is worth looking at before:

4.2.3 χ^2

$X_1, X_2, \ldots, X_n$ iid $\mathcal{N}(0,1)$-RV, then

$$\sum_{i=1}^{n} X_i^2$$

is by definition χ_n^2-distributed (say "Chisquare with n degrees of freedom, df). If $Y_1, \ldots, Y_n$ are iid $\mathcal{N}(\mu, \sigma^2)$-RV, then due to the Z-Transform

$$\sum_{i=1}^{n} \frac{(Y_i - \mu)^2}{\sigma^2}$$

is χ_n^2-distributed. Surprisingly: define

$$S^2 := \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

with $\overline{Y} := \sum_{i=1}^{n} Y_i/n$, then S^2/σ^2 has a χ_{n-1}^2-distribution. Rule of thumb: we loose a df per parameter that must be estimated (we estimated μ with $\overline{Y}$). Density of χ_n^2 is a monster and is omitted. $E[\chi_n^2] = n; V[\chi_n^2] = 2n$. Examples of the density (for $n > 5$):

4.2.4 F

Let U and V be independent, χ^2_m-, and χ^2_n-distributed RV's. Then

$$W := \frac{U/m}{V/n}$$

has an F-distribution with parameters m, n: $F_{m,n}$. Density is again a monster and is omitted. $E[W] = n/(n-2)$ (for $n > 2$). Examples of the density:

4.2.5 t

Let Y be $\mathcal{N}(0,1)$-distributed and Z be χ^2_n-distributed; Y independent of Z.

$$T_n := \frac{Y}{\sqrt{Z/n}}$$

has almost a $\mathcal{N}(0,1)$-distribution, but not quite, it is a Student-t-distribution with n df. Density is third monster and is omitted. We have $E[t_n] = 0$ (for $n > 1$), $V[t_n] = n/(n-2)$ for $n > 2$.

4.3 Exercises (more in course)

4.1 X a $\mathcal{N}(4, 25)$-RV. Find $P[-1 < X < 5]$.

4.2 X a $\mathcal{N}(20, 16)$-RV. Find $P[X > 16]$.

4.3 X a χ_3^2-RV. Find $P[X \geq 6.25]$.

4.4 Length of some animal is modelled with a normal distribution with $\mu = 3$ mm and $\sigma^2 = 2$. You look at 240 animals. How many of them do you expect to be less than 2.5 mm in length - how many between 2.8 and 3.1 mm?