Contents

Continuous Random Variables

All RVs up to now have been discrete

- Take on distinct values, e.g. in set $\{1, 2, 3\}$
- Often represent binary values or counts

What about continuous RVs?

- Take on real-values
- e.g. Travel time to work, temperature of this room, fraction of Irish population supporting Scotland in the rugb

Cumulative Distribution Function

Suppose Y is a random variable, which may be discrete or continuous valued.

- Recall $F_Y(y) := P(Y \leq y)$ is the cumulative distribution function (CDF)
- CDF exists and makes sense for both discrete and continuous valued random variables
- When Y takes discrete values $\{y_1, \dots, y_m\}$ then $F_Y(y) = \sum_{j:y_j \leq y} P(Y = y_j)$
- $F_Y(-\infty) = 0, F_Y(+\infty) = 1$
- Also
$$P(Y \leq b) = P(Y \leq a) + P(a < Y \leq b)$$
i.e.
$$F_Y(b) = F_Y(a) + P(a < Y \leq b)$$
therefore
$$P(a < Y \leq b) = F_Y(b) - F_Y(a)$$

CDF always starts at 0 and rises to 1, never decreasing.

Area Under a Curve

- Fit a series of rectangles under the curve, each of width h
- We know the area under a rectangle, its the height $\times$ width h
- Add up the areas of all the rectangles to get an estimate of the area under the curve
- As h gets smaller and smaller ($h \to 0$) this value becomes closer and closer to the true area
- Think of $f(y)dy$ as the area of the rectangle between y and $y + dy$ with dy infinitesimally small
- Write the area under curve between a and b as $\int_a^b f(y)dy$
- Think of integral as the sum of areas of rectangles each of width h as $h \to 0$
    - Integral symbol $\int$ is supposed to be suggestive of a sum
    - Can think of dy as h (infinitesimally small)

Example: CDF $F_Y(y)$ in right-hand plot is area under curve in left-hand plot between $-\infty$ and y, i.e. $F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt$

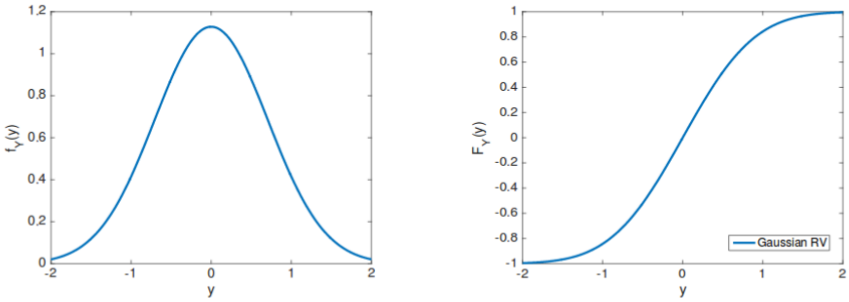

Figure 1: PDF and CDF

Continuous Random Variables: CDF and PDF

- For a continuous-valued variabled Y there exists a function $f_Y(y) \geq 0$ such that
$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt$$
- cf $F_Y(y) = \sum_{j:y_j \leq y} P(Y = y_j)$ in discrete-valued case
- f_Y is called the **probability density function** or **PDF** of Y
- $\int_{-\infty}^{\infty} f(y)dy = 1$ (since $\int_{-\infty}^{\infty} f(y)dy = F_Y(\infty) - P(Y \leq \infty) = 1$)

Note that tricky to define PDF f_Y for a discrete random variable since its CDF has "jumps" in it

- It follows that
$$P(a < Y \leq b) = F_Y(b) - F_Y(a) = \int_{-\infty}^{b} f_Y(t)dt - \int_{-\infty}^{a} f_Y(t)dt = \int_{a}^{b} f_Y(t)dt$$
- The probability density function $f(y)$ for random variable Y is *not* a probability, e.g. it can take values greater than 1
- Its the *area* under the PDF between points a and b that is the probability $P(a < Y \leq b)$
  - i.e. The total area under the curve is 1

Example: Uniform Random Variables

Y is a **uniform random variable** when it has PDF
$$f_Y(y) = \begin{cases} \frac{1}{\beta - \alpha} & \text{when } \alpha \leq y \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

- For $\alpha \leq a \leq b \leq \beta : P(a \leq Y \leq b) = \frac{b-a}{\beta - alpha}$
- `rand()` function in Matlab

A bus arrives at a stop every 10 minutes. You turn up at the stop at a time selected uniformly at random during the day and wait for 5 minutes. What is the probability that the bus turns up?

- Check the area under the PDF is 1. Area of left-hand trianles is $\frac{1}{2}$, area of the right hand traingle is the same. Total is 1.
- What is $P(0 \leq X \leq 1)$? Its the area under the PDF between points 0 and 1, i.e. the area of the right hand traingle, so $P(0 \leq X \leq 1) = 0.5$
- What is $P(0 \leq X \leq \infty)$? $f_X(x) = 0$ for $x > 1$, so $P(0 \leq X \leq \infty) = P(0 \leq X \leq 1) = 0.5$

Expectation and Variance

For dx infitesimally small,

$$P(x \leq X \leq x + dx) = F_X(x + dx) - F_X(x) \approx f_X(x)dx$$

so we can think of $f_X(x)dx$ as the probability that X takes a value between x and $x + dx$

For *discrete* RV X

- $E[X] = \sum_x xP(X = x)$
- $E[X^n] = \sum_x x^n P(X = x)$

For *continuous* RV X

- $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$
- $E_{X^n} = \int_{-\infty}^{\infty} x^n f_X(x)dx$

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

For both discrete and continuous random variables

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

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

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

(just replace sum with integral in previous proofs)

The Normal Distribution

Y is **Normal random variable** $Y \sim N(\mu, \sigma^2)$ when it has PDF

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

- $E[Y] = \mu, Var(Y) = \sigma^2$
- Symmetric about μ and defined for all real-valued x
- A Normal RV is also often called a **Gaussian random variable** and the Normal distribution referred to as the Gaussian distribution

Linearity of the Normal Distribution

Suppose $X \sim N(\mu, \sigma^2$. Let $Y = aX + b$, then

- $E[Y] = aE[X] + b - a\mu + b, Var(Y) = a^2 Var(X)$
- $Y \sim N(a\mu + b, a^2\sigma^2)$, i.e. Y is also Normally distributed

Suppose $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent RVs. Let $Z = X + Y$, then

- $E[Z] = E[X] + E[Y] = \mu_X + \mu_Y, Var(Z) = Var(X) + Var(Y) = \sigma_X^2 + \sigma_Y^2$
- $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$, i.e. W is also Normally distributed
- NB: Only holds for addition of Normal RVs, .e.g X^2 is not Normally distributed even if X is

Central Limit Theorem (CLT)

Why is it called the "Normal" distribution? Suggests its the "default". Coin toss example ahgain, but now we plot a distogram of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as N increases.

- Curve narrows as n increases, it concentrates as we already know from weak law of large numbers
- Curbe if roughly "bell-shaped" i.e. roughly Normal

Consider N indepdent and identically distributed random variables $X_1, \ldots, X_N$ each with mean μ and variance σ^2. Let $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$. Then

$$\bar{X} \sum N(\mu, \frac{\sigma^2}{N}) \text{ as } N \to \infty$$

- This says that as N increases the distrubion of $\bar{X}$ converges to a Normal (or Gaussian) distribution
- The distribution has mean μ and variance $\frac{\sigma^2}{N}$
- Variance $\frac{\sigma^2}{N} \to 0$ as $N \to \infty$
  - So distribution concentrates around the mean μ as N increases

Confidence Intervals (Again)

- Recall that when a random variable lies in an interval $a \leq X \leq b$ with a specified probability we call this a confidence interval
- When $X \sim N(\mu, \sigma^2)$:

$$P(-\sigma \leq X - \mu \leq \sigma) \approx 0.68$$

$$P(-2\sigma \leq X - \mu \leq 2\sigma) \approx 0.95$$

$$P(-3\sigma \leq X - \mu \leq 3\sigma) \approx 0.997$$

- There are $1\sigma, 2\sigma, 3\sigma$ confidence intervals
- $\mu \pm 2\sigma$ in the 95 confidence interval for a Normal random variable with mean μ and variance σ^2
  - In practice often use either $\mu \pm \sigma$ or $\mu \pm 3\sigma$ as confidence intervals
- Recall claim by Goldman Sachs that crash was a 25σ event

But

- These confidence intervals differ from those we previously derived from Chebyshev and Chernoff inequalities
  - Chebyshev and Chernoff confidence intervals are actual confidence intervals
  - Those derived from CLT are only approximate (accuracy depends on how large N is)
- We need to be careful to check that N is large enough that distribution really is almost Gaussian
  - This might need large N

Example: Running Time of New Algorithm

Suppose we have an algorithm to test. We run it N times and measure the time to complete, gives measurements $X_1, \ldots, X_N$

- Mean running time is $\mu = 1$, variance is $\sigma^2 = 4$
- How many trials do we need to make so that the measured sample mean running time is within 0.5s of the mean μ with 95% probability?
  - $P(\mid X - \mu \mid \geq 0.5) \leq 0.05$ where $X = \frac{1}{N}\sum_{k=1}^{N} X_k$
- CLT tells us that $X \sim N(\mu, \frac{sigma^2}{N})$ for large N
- So we need $2\sigma = 2\sqrt{\frac{\sigma^2}{N}} = 0.5$ i.e. $N \geq 64$

Wrap Up

We have three different approaches for estimating confidence intervals, each with their pros and cons

- CLT: $\bar{X} \sim N(\mu, \frac{\sigma^2}{N}$ as $N \to \infty$
  - Gives full distribution of $\bar{X}$
  - Only requires mean and variance to fully describe this distribution
  - But is an approximation when N finite, and hard to be sure how accurate (how big should N be?)
- Chebyshev and Chernoff
  - Provides an actual bound
  - Works for all N
  - But loose in general
- Bootstrapping
  - Gives full distribution of $\bar{X}$, doesn't assume Normality
  - But is an approximation where N finite, and hard to be sure how accurate (how big should N be?)
  - Requires availability of all N measurements