CS 70 Discrete Mathematics and Probability Theory
Summer 2025 Tate

DIS 06B

Continuous Probability Intro II

Normal (Gaussian) Distribution: $X \sim N(\mu, \sigma^2)$

The normal distribution occurs frequently in nature, mostly due to the Central Limit Theorem.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$F_X(x) = \Phi(x)$$

Note that there is no closed form expression for the CDF of the normal distribution.

Properties:

- A **standard normal** distribution is denoted as $Z \sim N(0,1)$
- If $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$
- Generally, if $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent, then
$$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2).$$

Central Limit Theorem: Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with mean μ and variance σ^2, and let

$$S_n = \sum_{i=1}^{n} X_i \qquad A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Note that

$$\mathbb{E}[A_n] = \mu \qquad \text{Var}(A_n) = \frac{\sigma^2}{n}$$

The central limit theorem states that as $n \to \infty$, $A_n \to N(\mu, \frac{\sigma^2}{n})$. Or,

$$S_n \to N(n\mu, n\sigma^2)$$
$$\frac{A_n - \mu}{\sigma/\sqrt{n}} \to N(0,1)$$
$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to N(0,1)$$

These four equations are all equivalent formations of the same idea, which is that the sample mean (of i.i.d random variables) will converge to a normal distribution preserving the mean and variance of the sample mean, as $n \to \infty$ (and the same holds for the sample sum, and shifted/scaled versions of them).

1 Sum of Independent Gaussians

Note 21

In this question, we will introduce an important property of the Gaussian distribution: the sum of independent Gaussians is also a Gaussian.

Let X and Y be independent standard Gaussian random variables.

(a) What is the joint density of X and Y?

(b) Observe that the joint density of X and Y, $f_{X,Y}(x,y)$, only depends on the quantity x^2+y^2, which is the distance from the origin. In other words, the Gaussian is *rotationally symmetric*. Next, we will try to find the density of $X+Y$. To do this, draw a picture of the Cartesian plane and draw the region $x+y \leq c$, where c is a real number of your choice.

(c) Now, rotate your picture clockwise by $\pi/4$ so that the line $X+Y=c$ is now vertical. Redraw your figure. Let X' and Y' denote the random variables which correspond to the $\pi/4$ clockwise rotation of (X,Y). Express the new shaded region in terms of X' and Y'.

(d) By rotational symmetry of the Gaussian, (X',Y') has the same distribution as (X,Y). Argue that $X+Y$ has the same distribution as $\sqrt{2}Z$, where Z is a standard Gaussian.

This proves the following important fact: *the sum of independent Gaussians is also a Gaussian*. Notice that $X \sim N(0,1)$, $Y \sim N(0,1)$ and $X+Y \sim N(0,2)$. *In general, if X and Y are independent Gaussians, then $X+Y$ is a Gaussian with mean $\mu_X+\mu_Y$ and variance $\sigma_X^2+\sigma_Y^2$.*

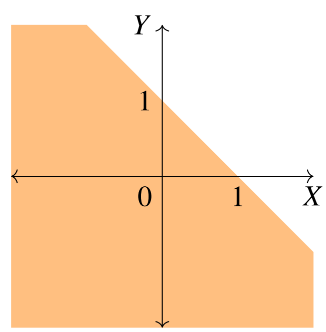
(e) Prove that for $n \geq 1$, if $X_1, X_2, \cdots, X_n$ are i.i.d standard normal, then $S_n = \sum_{i=1}^n X_i$ is normal with mean 0 and variance n.

Solution:

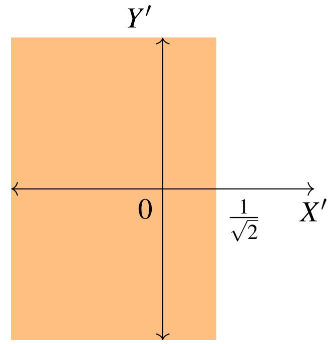
(a) Because X,Y are independent, we have

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right).$$

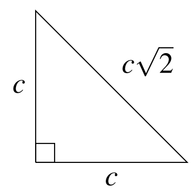
(b) We draw the line for $c=1$.



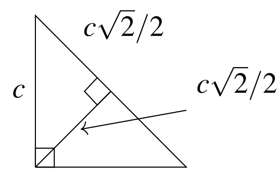
(c) Here is the new figure after the rotation (for $c=1$).



For general $c \in \mathbb{R}$, the new region is $\{X' \le c/\sqrt{2}\}$. To see why, draw the triangle: We want



to find the distance between the origin and the long side of the triangle, and we can do so by adding a diagonal:



(d) We observe that $\mathbb{P}(X+Y \le c) = \mathbb{P}(X' \le c/\sqrt{2}) = \mathbb{P}(\sqrt{2}X' \le c)$, where X' is a standard Gaussian by rotational symmetry, so this proves the claim.

(e) By induction on n.

Base Case: $n = 1$. This is trivially true, as $S_1 = X_1 \sim N(0,1)$.

Inductive Hypothesis: Assume that for some $n = k$, $S_k \sim N(0,k)$.

Inductive Step: We'll show that for $n = k+1$, $S_{k+1} \sim N(0,k+1)$.

Notice that we can rewrite $S_{k+1} = \sum_{i=1}^{k+1} X_i = \sum_{i=1}^{k} X_i + X_{k+1} = S_k + X_{k+1}$. By our inductive hypothesis, S_k is normal with mean 0 and variance k. And X_{k+1} is normal with mean 0 and variance 1. And since the X_i are independent, S_k and X_{k+1} are also independent. Applying the property in (d) gives us that S_{k+1} is normal with mean 0 and variance $k+1$.

2 Binomial Concentration

Note 21

Here, we will prove that the binomial distribution is *concentrated* about its mean as the number of trials tends to ∞. Suppose we have i.i.d. trials, each with a probability of success $1/2$. Let S_n be the number of successes in the first n trials (n is a positive integer).

(a) Compute the mean and variance of S_n.

(b) Define Z_n in terms of S_n such that Z_n has mean 0 and variance 1.

(c) What is the distribution of Z_n as $n \to \infty$?

(d) Use the bound $\mathbb{P}[Z > z] \le (\sqrt{2\pi}z)^{-1}\mathrm{e}^{-z^2/2}$ (where Z is a standard normal) to approximately bound $\mathbb{P}[S_n/n > 1/2+\delta]$, where $\delta > 0$.

Solution:

(a) Since $S_n \sim \text{Binomial}(n, \frac{1}{2})$, we have $\mathbb{E}[S_n] = \frac{n}{2}$ and $\text{Var}(S_n) = \frac{n}{4}$.

(b) We can define

$$Z_n := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n/2}{\sqrt{n}/2}.$$

In particular, we subtract the mean and divide by the standard deviation to normalize S_n.

To check, we have

$$\begin{aligned}\mathbb{E}[Z_n] &= \frac{1}{\sqrt{n}/2}\mathbb{E}\left[S_n - \frac{n}{2}\right] = \frac{1}{\sqrt{n}/2}\left(\mathbb{E}[S_n] - \frac{n}{2}\right) = 0,\\ \text{Var}(Z_n) &= \frac{1}{n/4}\text{Var}\left(S_n - \frac{n}{2}\right) = \frac{1}{n/4}\text{Var}(S_n) = 1,\end{aligned}$$

since $S_n \sim \text{Binomial}(n, 1/2)$.

(c) The central limit theorem tells us that $Z_n \to N(0,1)$.

(d) In order to apply the bound, we must apply it to Z_n.

$$\begin{aligned}\mathbb{P}\left[\frac{S_n}{n} > \frac{1}{2}+\delta\right] &= \mathbb{P}\left[\frac{S_n - n/2}{n} > \delta\right] = \mathbb{P}\left[\frac{S_n - n/2}{\sqrt{n}/2} > 2\delta\sqrt{n}\right] \approx \mathbb{P}[Z_n > 2\delta\sqrt{n}]\\ &\le \frac{1}{2^{3/2}\delta\sqrt{\pi n}}e^{-2\delta^2 n}\end{aligned}$$

3 Erasures, Bounds, and Probabilities

Note 21

Alice is sending 1000 bits to Bob. The probability that a bit gets erased is p, and the erasure of each bit is independent of the others.

Alice is using a scheme that can tolerate up to one-fifth of the bits being erased. That is, as long as Bob receives at least 801 of the 1000 bits correctly, he can decode Alice's message.

In other words, Bob becomes unable to decode Alice's message only if 200 or more bits are erased. We call this a "communication breakdown", and we want the probability of a communication breakdown to be at most 10^{-6}.

(a) Use Chebyshev's inequality to upper bound p such that the probability of a communications breakdown is at most 10^{-6}.

(b) As the CLT would suggest, approximate the fraction of erasures by a Gaussian random variable (with suitable mean and variance). Use this to find an approximate bound for p such that the probability of a communications breakdown is at most 10^{-6}.

You may use that $\Phi^{-1}(1-10^{-6}) \approx 4.753$.

Solution:

(a) Let X be the random variable denoting the number of erasures. Chebyshev's inequality states the following:

$$\mathbb{P}[|X-\mu_X| \geq k] \leq \frac{\sigma_X^2}{k^2}.$$

This gives us the bound

$$\begin{aligned}\mathbb{P}[X \geq 200] &= \mathbb{P}[X-\mu_X \geq 200-\mu_X] \\ &\leq \mathbb{P}[|X-\mu_X| \geq 200-\mu_X] \\ &\leq \frac{\sigma_X^2}{(200-\mu_X)^2}\end{aligned}$$

Since $X \sim \text{Binomial}(1000, p)$, we have $\mu_X = 1000p$ and $\sigma_X^2 = 1000p(1-p)$. Substituting these values in, we have

$$\mathbb{P}[X \geq 200] \leq \frac{1000p(1-p)}{(200-1000p)^2} = \frac{p(1-p)}{40(1-5p)^2}.$$

To meet our objective, we just have to ensure that

$$\mathbb{P}[X \geq 200] \leq \frac{p(1-p)}{40(1-5p)^2} \leq 10^{-6},$$

which yields an upper bound of about 3.998×10^{-5} for p.

(b) Let Y be equal to the fraction of erasures, i.e. $\frac{X}{1000}$. Using properties of expectation and variance, we can see that

$$\begin{aligned}\mathbb{E}[Y] &= p \\ \text{Var}(Y) &= \text{Var}(X) \cdot \frac{1}{1000^2} = \frac{p(1-p)}{1000}\end{aligned}$$

Therefore, by Central Limit Theorem, we can say that Y is roughly a normal distribution with that mean and variance. Since we are interested in the event that $Y \geq 0.2$, let's figure out how many standard deviations above the mean 0.2 is:

$$\frac{0.2-p}{\sqrt{\frac{p(1-p)}{1000}}} = \frac{(0.2-p)\sqrt{1000}}{\sqrt{p(1-p)}}.$$

Therefore, the probability that we get a failure should be approximately (by CLT),

$$1 - \Phi\left(\frac{(0.2-p)\sqrt{1000}}{\sqrt{p(1-p)}}\right)$$

where Φ is the CDF of a standard normal variable. Setting this to be at most 10^{-6} gives us

$$\Phi\left(\frac{(0.2-p)\sqrt{1000}}{\sqrt{p(1-p)}}\right) \geq 1 - 10^{-6}$$

And, since $\Phi^{-1}(1-10^{-6}) \approx 4.753$, we solve the inequality

$$\frac{(0.2-p)\sqrt{1000}}{\sqrt{p(1-p)}} \geq 4.753$$

This yields that we need $p \leq 0.1468$.

Note that this gives quite a different value from the previous parts. This is because the Central Limit Theorem gives a much tighter approximation for tail events than Markov's and Chebyshev's. However, we can only apply the Central Limit Theorem because n is large.

Therefore, we do not need p to be so low to achieve a communication breakdown probability of 10^{-6}. The other bounds required us to need a probability of on the order of 10^{-5}, but here we realize that we only need it to be less than 0.1468. (The true bound is .1459.) Quite drastic!