CS 70 Discrete Mathematics and Probability Theory
Fall 2024 Hug, Rao

DIS 12B

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

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

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

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

1 Interesting Gaussians

Note 21

(a) If $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$ are independent, then what is $\mathbb{E}\left[(X+Y)^k\right]$ for any *odd* $k \in \mathbb{N}$?

(b) Let $f_{\mu,\sigma}(x)$ be the density of a $N(\mu, \sigma^2)$ random variable, and let X be distributed according to $\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha) f_{\mu_2,\sigma_2}(x)$ for some $\alpha \in [0,1]$. Compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$. Is X normally distributed?

Solution:

(a) $\mathbb{E}\left[(X+Y)^k\right] = 0$.

Since X and Y are Gaussians, so must $Z = X + Y$ be. Specifically, $Z \sim N(0, \sigma_X^2 + \sigma_Y^2)$. Thus, the PDF f_Z of Z is still symmetric about the origin; that is, it is an even function, i.e. $f_Z(x) = f_Z(-x)$ for any $a, b \in \mathbb{R}$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[(X+Y)^k\right] = \mathbb{E}\left[Z^k\right] &= \int_{-\infty}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{0} x^k f_Z(x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= \int_{0}^{\infty} (-x)^k f_Z(-x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= -\int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= 0,
\end{aligned}$$

since k is odd.

Note that we could've just concluded that $\int_{-\infty}^{\infty} x^k f_Z(x)\,\mathrm{d}x = 0$ due to the fact that $x^k f_Z(x)$ is an odd function (since x^k is an odd function for odd k), and the integral from $(-a, a)$ for any odd function will evalute to 0.

Also note that adding two RVs is NOT equivalent to adding their PDFs. Instead, adding two RVs is equivalent to convolving their PDFs. As an example, for random variables $X + Y = Z$, it is true that $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx$.

(b) $\mathbb{E}[X] = \alpha\mu_1 + (1-\alpha)\mu_2$, $\mathrm{Var}(X) = \alpha\left(\sigma_1^2 + \mu_1^2\right) + (1-\alpha)\left(\sigma_2^2 + \mu_2^2\right) - (\mathbb{E}[X])^2$. No, X is not necessarily normally distributed.

$$\begin{aligned}
\mathbb{E}[X] := \mu &= \int_{-\infty}^{\infty} x\left(\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha) f_{\mu_2,\sigma_2}(x)\right)\mathrm{d}x \\
&= \alpha \int_{-\infty}^{\infty} x f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} x f_{\mu_2,\sigma_2}(x)\,\mathrm{d}x = \alpha\mu_1 + (1-\alpha)\mu_2 \\
\mathrm{Var}(X) := \sigma^2 = \mathbb{E}\left[X^2\right] - \mu^2 &= \alpha \int_{-\infty}^{\infty} x^2 f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} x^2 f_{\mu_2,\sigma_2}(x)\,\mathrm{d}x - \mu^2 \\
&= \alpha\left(\sigma_1^2 + \mu_1^2\right) + (1-\alpha)\left(\sigma_2^2 + \mu_2^2\right) - \mu^2.
\end{aligned}$$

We know that the density of $N(\mu,\sigma)$ has a unique maximum at $x=\mu$; however, if, e.g. $\alpha=1/2, \mu_1=-10, \mu_2=10, \sigma_1=\sigma_2=1$, then $\alpha f_{\mu_1,\sigma_1}+(1-\alpha)f_{\mu_2,\sigma_2}$ has two maxima, and so cannot be the density of a Gaussian.

Explanation of integrals: $\int_{-\infty}^{\infty} x f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x$ becomes $\mathbb{E}[X_1]$ for X_1 with PDF $f_{\mu_1,\sigma_1}(x)$, which is μ_1 by definition.

$\int_{-\infty}^{\infty} x^2 f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x$ becomes $\mathbb{E}\left[X_1^2\right]$ for X_1 with PDF $f_{\mu_1,\sigma_1}(x)$. $\mathbb{E}\left[X_1^2\right]=\mathrm{Var}(X_1)+\mathbb{E}[X_1]^2=\sigma_1^2+\mu_1^2$ by definition.

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) How should we define Z_n in terms of S_n to ensure 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 $\mathbb{P}[Z>z]\le(\sqrt{2\pi}z)^{-1}\mathrm{e}^{-z^2/2}$ when Z is a standard normal in order to approximately bound $\mathbb{P}[S_n/n>1/2+\delta]$, where $\delta>0$.

Solution:

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

(b) We can define

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

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

To check, we have

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

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

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

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

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

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

$$\mathbb{E}[Y] = p$$
$$\mathrm{Var}(Y) = \mathrm{Var}(X) \cdot \frac{1}{1000^2} = \frac{p(1-p)}{1000}$$

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!