Class Notes 2-27-2019

- **The Gaussian Distribution:**
A random variable Z has the standard normal distribution (or standard Gaussian distribution) and we write $Z \sim \mathcal{N}(0,1)$ if Z has pdf

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

Its cdf is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2}\, ds$$

There is no closed form anti-derivative for $\varphi(x)$. To show that it is a density we resort to the trick of computing

$$\left(\int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\, dy\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy$$

Rather than summing the areas $dx\ dy$ of the rectangles $[x, x+dx] \times [y, y+dy]$ multiplied by the function value $\exp(-(x^2+y^2)/2)$ at its rectangle corner (x,y) we represent the points (x,y) in the plane $\mathbb{R}^2$ by polar coordinates, namely $x = r\cos\theta$ and $y = r\sin\theta$, and sum the wedge shaped areas cut out of the annulus defined by radii r and $r+dr$ and by angles θ and $\theta + d\theta$ and multiply them by the function value $\exp(-(x^2+y^2)/2) = \exp(r^2/2)$ at the wedge corner (r,θ). The annulus area is $\pi(r+dr)^2 - \pi r^2$ and the angular $d\theta$ slices represent $\frac{d\theta}{2\pi}$ of the full circumference 2π. Thus these wedge areas are

$$(\pi(r+dr)^2 - \pi r^2)\frac{d\theta}{2\pi} = r dr d\theta + \frac{1}{2} dr^2 d\theta \approx r dr d\theta$$

Thus we get with $x^2 + y^2 = r^2\cos^2\theta + r^2\sin^2\theta = r^2$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy = \int_0^{2\pi}\int_0^{\infty} r e^{-r^2/2} dr d\theta = 2\pi\left(-e^{-r^2/2}\right)\Big|_0^{\infty} = 2\pi \Rightarrow \int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

i.e., $\varphi(x)$ is a density.

A more compact way of showing the same is to turn the 2 dimensional integral into a 1 dimensional one by realizing that $e^{-(x^2+y^2)/2}$ is constant on concentric circles around $(0,0)$. Rather than integrating/summing over small rectangular cells we integrate/sum over thin annuli of radii r and $r+dr$. These have area $\pi(r+dr)^2 - \pi r^2 = 2\pi r dr + \pi(dr)^2 \approx 2\pi r$. On each such annulus the function $e^{-(x^2+y^2)/2} \approx e^{-r^2/2}$ and thus the 2 dimensional integral becomes

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy = \int_0^{\infty} 2\pi r e^{-r^2/2} dr = -2\pi e^{-r^2/2}\Big|_0^{\infty} = 2\pi$$

- **The cdf $\Phi(x)$:**
Appendix E gives us $\Phi(x)$ as tabled values for $x \in [0, 3.49]$. Because of the symmetry of $\varphi(x)$ around zero, i.e., $\varphi(-x) = \varphi(x)$, we get $P(Z \le -x) = \Phi(-x) = 1 - \Phi(x) = P(Z \ge x)$ and thus we also get $\Phi(x)$ for $x \in [-3.49, 0]$.

- **Examples of Standard Normal Probabilities:**

$$\begin{aligned} P(Z \le 1.7) &= \Phi(1.7) = .9554 \\ P(Z \le -1.7) &= 1 - \Phi(1.7) = .0446 \\ P(-1 \le Z \le 1.5) &= P(Z \le 1.5) - P(Z \le -1) \\ &= \Phi(1.5) - (1 - \Phi(1)) = .9332 - (1 - .8413) \end{aligned}$$

Find z such that $P(-z < Z < z) = 2/3$, i.e., $P(Z < z) = 5/6 = .8333$.
We have $P(Z \le .96) = .8315$ and $P(Z \le .97) = .8340$, thus $z = .97$ comes close. People tend to be more cavalier and use $z = 1$ instead to store $P(-1 < Z < 1) \approx 2/3$ in their memory bank, when in fact $P(-1 < Z < 1) = 0.6827 > .6666$.

- **Mean and Variance of the Standard Normal Z:**

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2}\, dx$$

is an improper integral. Need to show that it exists no matter how we approach $\pm\infty$. To that end we examine

$$\int_{-\infty}^{\infty} |x| e^{-x^2/2}\, dx = 2 \int_0^{\infty} x e^{-x^2/2}\, dx = -2e^{-x^2/2}\Big|_0^{\infty} = 2 < \infty$$

thus $E(Z)$ exists and is zero because $g(x) = x\varphi(x)$ is odd, i.e., $g(-x) = -g(x)$, so that

$$\int_{-a}^{a} g(x)dx = 0 \quad \text{for all } a > 0 \text{ and thus} \quad \lim_{a\to\infty} \int_{-a}^{a} g(x)dx = \int_{-\infty}^{\infty} g(x)dx = 0 = E(X)$$

$$\begin{aligned} E(Z^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx \quad \text{integrating by parts} \\ &= -\frac{1}{\sqrt{2\pi}}\, x e^{-x^2/2}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx = 1 \quad \text{thus } \mathrm{var}(Z) = 1 \end{aligned}$$

where we used $\lim_{x\to\pm\infty} x e^{-x^2/2} = 0$.

- **General Normal Random Variables:**
General normal random variables can be viewed as linear transforms of a standard normal random variable Z. Let $\mu \in \mathbb{R}$ and $\sigma > 0$ then $X = \sigma Z + \mu$ is said to have a normal distribution with mean μ and variance σ^2 and we write $X \sim \mathcal{N}(\mu, \sigma^2)$. Its cdf is

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

and by differentiation its density is

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

If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$ with $a \ne 0$, then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$,
since $Y = aX + b = a(\sigma Z + \mu) + b = a\sigma Z + a\mu + b$. In particular, $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

- **Normal Probability Calculations:**
  Say we have $X \sim \mathcal{N}(-3, 4)$, find $P(X \leq -1.7)$.

$$P(X \leq -1.7) = P\left(\frac{X-(-3)}{2} \leq \frac{-1.7-(-3)}{2}\right) = P(Z \leq .65) = \Phi(.65) = .7422$$

  If $X \sim \mathcal{N}(\mu, \sigma^2)$, find $P(|X - \mu| > 2\sigma)$.

$$P(|X-\mu| > 2\sigma) = P(|(X-\mu)/\sigma| > 2) = P(|Z| > 2) = 2(1-\Phi(2)) \approx 2(1-.9772) = .0456$$

  roughly 95% of the probability under the normal curve lies within $\mu \pm 2\sigma$
  and about 66% lies within $\mu \pm \sigma$.