Lecture 14, Feb 10, 2023

The Normal (Gaussian) Distribution

Definition

Given a mean μ and variance σ^2, the *normal distribution* is given by

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

- This gives a symmetric bell centered around the mean μ, with width proportional to σ
  - μ is a translation
  - σ gets larger as the curve gets wider and flatter
  - $\lim_{\sigma\to 0} n(x;\mu,\sigma) = \delta(x-\mu)$
- The Gaussian is important due to the central limit theorem: taking a large number of random variables and taking their average, it will give the normal distribution regardless of the distribution of the individual random variables
- $$\begin{aligned}\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x\right)^2 &= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x\int_{-\infty}^{\infty}e^{-\frac{y^2}{2\sigma^2}}\,\mathrm{d}y \\ &= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{x^2+y^2}{2\sigma^2}}\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2\pi\sigma^2}\int_0^{2\pi}\int_0^{\infty}e^{-\frac{r^2}{2\sigma^2}}r\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \frac{1}{4\pi\sigma^2}\int_0^{2\pi}\int_0^{\infty}e^{-\frac{s}{2\sigma^2}}\,\mathrm{d}s\,\mathrm{d}\theta \\ &= \frac{1}{2\sigma^2}\int_0^{\infty}e^{-\frac{s}{2\sigma^2}}\,\mathrm{d}s \\ &= \frac{1}{2\sigma^2}\left[-2\sigma^2e^{-\frac{s}{2\sigma^2}}\right]_0^{\infty} \\ &= 1\end{aligned}$$
- $$\begin{aligned}E[X] &= \int_{-\infty}^{\infty}x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma z+\mu)e^{-\frac{z^2}{2}}\sigma\,\mathrm{d}z \\ &= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ze^{-\frac{z^2}{2}}\,\mathrm{d}z + \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}}\,\mathrm{d}z \\ &= 0 + \mu\int_{-\infty}^{\infty}n(z;0,1)\,\mathrm{d}z \\ &= \mu\end{aligned}$$
  - Substitute $z = \frac{x-\mu}{\sigma}$
  - $\int_{-\infty}^{\infty}ze^{-\frac{z^2}{2}}\,\mathrm{d}z = 0$ because the integrand is an odd function
- Using a similar argument we may show that the variance is σ^2

Definition

$n(x;0,1)$ is referred to as the *standard normal distribution*

$$\Phi(x) = \int_{-\infty}^{x} n(y;0,1)\,\mathrm{d}y$$

is the cumulative distribution function of the standard normal, so

$$P(A \leq X \leq B) = \Phi(B) - \Phi(A)$$

- Note Φ is not analytically evaluable, so there are usually tables of values for it
- Suppose X has PDF $n(x;\mu,\sigma)$; let $Z = \dfrac{X-\mu}{\sigma}$, then Z has PDF $n(x;0,1)$, which is the standard normal
  - $P(X \leq x) = \int_{-\infty}^{x} n(x;\mu,\sigma)\,\mathrm{d}x$
    $= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{\sigma^2}}\,\mathrm{d}t$
    $= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{s^2}{2}}\sigma\,\mathrm{d}s$
    $= \int_{-\infty}^{\frac{x-\mu}{\sigma}} n(s;0,1)\,\mathrm{d}s$
    $= P\left(Z \leq \frac{x-\mu}{\sigma}\right)$
    $= \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Therefore $P(A \leq X \leq B) = \Phi\left(\frac{A-\mu}{\sigma}\right) - \Phi\left(\frac{B-\mu}{\sigma}\right)$
- Example: Suppose X is a random variable with distribution $n(x;5,2)$; find $P(-1 \leq X \leq 4)$
  - Need to transform this into the standard normal
  - Let $Z = \dfrac{X-5}{2}$ then Z has the standard normal distribution and CDF Φ
  - $P(-1 \leq X \leq 4) = P\left(\frac{-1-5}{2} \leq \frac{X-5}{2} \leq \frac{4-5}{2}\right) = P\left(-3 \leq Z \leq -\frac{1}{2}\right) = \Phi\left(-\frac{1}{2}\right) - \Phi(-3) = 0.3072$
  - $\Phi(x)$ is `normcdf` in MATLAB