**Normal Distributions**

* **Normal (Gaussian) distribution**
  + R. v. X has a normal distribution with parameters μ and σ − i.e. X ~ N(μ, σ) − if its PDF is given by
    - f(x) = f(x, μ, σ) =
    - Where μ = mean and σ = standard deviation
    - Notation can also be X ~ N(μ, σ2) – mean & variance as parameters
    - E(X) = μ
    - Var(X) = σ2
  + If X ~ N(μ1, σ1) and Y ~ N(μ2, σ2)
    - Then X + Y ~ N(μ1 + μ2, √(σ12 + σ22))
    - Any linear combination: aX + bY + c ~ N(aμ1 + bμ2 + c, √(a2σ12 + b2σ22))
  + The PDF f(x) is not integrable, hence the normal distribution table is used
  + **Standard normal distribution**
    - If X ~ N(μ, σ), Z = (X − μ)/σ ~ N(0, 1)
    - P(a < Z < b) = F(b) – F(a) where F(x) is obtained from the table
    - P(a < Z) = 1 – F(a)
    - P(Z < b) = F(b)
    - Because the PDF is symmetric, P(Z > z) = P(Z < -z) for any z ∈ R
    - P(a < X < b) = P((a − μ)/σ < Z < (b − μ)/σ)
  + Ex: bolts produced have diameters Y ~ N(10, 1)
    - Washers produced have inner diameters X ~ (11, 0.5)
    - P(bolt fits inside washer) = P(X > Y) = P(X – Y > 0) = ?
    - X – Y ~ N(11 – 10, √(0.52 + (-1)212)) ~ N(1, 1.118)
    - Z = (X – Y – 1)/1.118 ~ N(0, 1)
    - P(X – Y > 0) = P((X – Y – 1)/1.118 > (0 – 1)/1.118))

= P(Z > -0.89)

= P(Z < 0.89) by symmetry

= 0.8133 from table

* + α = P(Z > z\_α)
    - i.e. z\_α is the z-score such that the probability of the value of the r. v. exceeding it is α
* **Central Limit Theorem**
  + Independent random variables – they have no influence on each other’s values
    - Given discrete r. v. X & Y, their joint probability function is
      * f(x, y) = P(X = x ∩ Y = y)
    - X & Y are independent if
      * f(x, y) = P(X = x) ⋅ P (Y = y) = fx(x)fy(y) for all x, y
  + Consider independent & identically distributed r. v. X1 … Xn all having:
    - E(Xi) = μ and Var(Xi) = σ2
  + **CLT – Sum:**
    - As n → ∞, S = X1 + … + Xn → ~ N(nμ, σ√n)
    - E(S) = nμ
    - Var(S) = nσ2
  + **CLT – average:**
    - As n → ∞, X-bar = (X1 + … + Xn)/n → ~ N(µ, σ/√n)
    - E(X-bar) = μ
    - Var(X-bar) = σ2/n
  + In general, CLT can be applied for n ≥ 30
  + If Xi are normally distributed, S and X-bar are exactly normal distributions
  + If Xi are not normally distributed, S and X-bar are approximately normal distributions
* **Continuity correction**
  + X ~ Bin(n, p) can be thought of as X = ∑(1 → n) Xi where Xi is a binary variable
    - CLT states that as n → ∞, X → a normal distr.
  + Continuity correction – improves the approximation of discrete probabilities
    - i.e. approximate the probability histogram using a continuous probability distribution with CDF F(x)
    - For a discrete distribution with width = w, the value k is represented by the interval k – w/2 and k + w/2
      * E.g. 7 people → (6.5, 7.5)
      * E.g. at least 3 people → (2.5, ∞)
  + For Bin. distribution, use μ = np and σ2 = npq
    - i.e. X ~ Bin(n, p) ~ (approx.) N(np, √(npq))
  + Ex: histogram on 0 ≤ X ≤ 10 with bar width = 1 and integer value centered in the bar
    - Approximate by integrating on (-0.5, 10.5) instead of (0, 10)
    - Suppose X ~ Bin(10, ½)
    - Then P(X=0) = 1/1024, P(X=1) = 10/1024, P(X=2) = 45/1024, …
    - μ = np = 5
    - σ2 = np(1 – p) = 2.5
    - Approximate X by Y ~ N(5, √2.5)
    - Z = (Y – 5)/√2.5 ~ N(0, 1)