

STAT 3011 Discussion 015

Week 5

Talha Hamza

University of Minnesota

Fall 2025

Recap: Normal Distribution

- The **normal distribution** is a continuous probability distribution characterized by:
 - **Mean** (μ): The center of the distribution.
 - **Standard deviation** (σ): Measures data spread.
- The **standard normal distribution** is a special case where:
 - $\mu = 0, \sigma = 1$
 - Denoted as: $Z \sim N(0, 1)$

Z-Score Transformation

Any normal distribution $X \sim N(\mu, \sigma)$ can be converted to the standard normal distribution using the **z-score**:

$$Z = \frac{X - \mu}{\sigma}$$

- This transformation allows us to use standard normal tables or R's `pnorm()` function for probability calculations.
- **Key Idea:** The z-score tells us how far a value is from the mean, making different normal distributions comparable.

Using `pnorm()` in R

To find the probability under the standard normal curve:

- $P(Z < z)$: `pnorm(z)`
- $P(Z > z)$: `1 - pnorm(z)`
- $P(z_a < Z < z_b)$: `pnorm(z_b) - pnorm(z_a)`

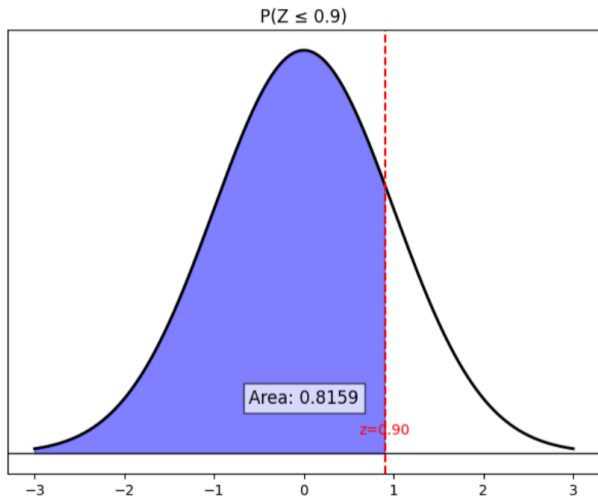
Note: `pnorm()` can also be used for any normal distribution $X \sim N(\mu, \sigma)$ by specifying the mean and standard deviation:

- $P(X < x)$: `pnorm(x, mean = μ , sd = σ)`
- $P(X > x)$: `1 - pnorm(x, mean = μ , sd = σ)`

Examples

Example 1: Find $P(Z < 0.9)$

```
pnorm(0.9)  
# Output: 0.8159
```



Using qnorm() in R

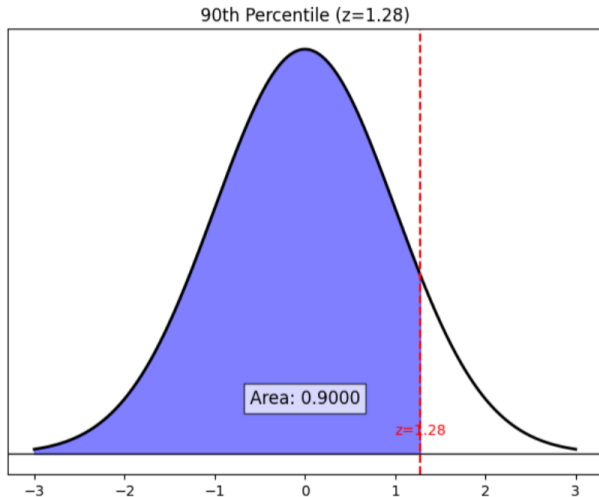
To find the z-value corresponding to a given percentile:

- z such that $P(Z < z) = p$: 'qnorm(p)'
- z such that $P(Z > z) = p$: 'qnorm(1 - p)'

Example

Example 4: Find the z-value that marks the 90th percentile of the standard normal distribution.

```
qnorm(0.9)  
# Output: 1.2816
```



`pnorm()` vs. `qnorm()`

- **`pnorm(z)`** → Returns the probability (area under the curve) below z .
"Given a z-score, what proportion of values are smaller?"
- **`qnorm(p)`** → Returns the z-score that corresponds to a given probability.
"Given a probability, what z-score separates the lower area?"

Conclusion: `pnorm()` finds the area given a z-score, `qnorm()` finds the z-score given the area.

Q-Q Plots

Q-Q (Quantile-Quantile) plots help us assess if data follows a normal distribution and reveal skewness patterns.

Interpreting the patterns:

- **Points follow straight line:** Data is normally distributed
- **Curve upward (right skew):** More extreme large values
- **Curve downward (left skew):** More extreme small values

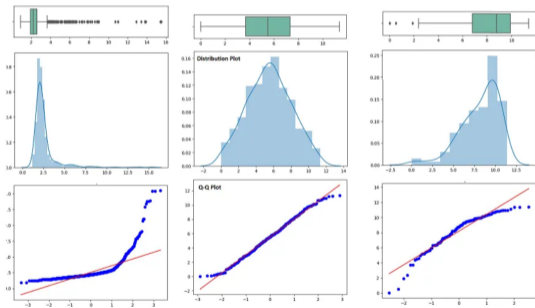


Image credits: <https://amitprius.medium.com/fully-understand-q-q-plot-for-probability-distribution-in-machine-learning-7ba16166cae6>

Introduction to Binomial Distribution

Binomial Distribution describes the number of successes in a fixed number of independent trials.

- **Fixed number of trials** (n): Known in advance
- **Independent trials**: Outcome of one doesn't affect others
- **Constant probability** (p): Same probability of success for each trial
- **Binary outcomes**: Each trial results in success or failure

Notation: $X \sim \text{Binomial}(n, p)$

Binomial Distribution Formula

The probability of getting exactly k successes in n trials is:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

where:

- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient
- p = probability of success on each trial
- n = number of trials
- k = number of successes (0, 1, 2, ..., n)

Binomial Distribution

- **Success** is simply the outcome we're counting
- It doesn't imply "good" or "desirable" in conventional terms
- A "success" could be:
 - A defective item in quality control
 - A patient experiencing side effects
 - A loan default in financial risk analysis
 - A machine breakdown in reliability studies

Key properties:

- Mean: $\mu = np$
- Variance: $\sigma^2 = np(1 - p)$
- Standard deviation: $\sigma = \sqrt{np(1 - p)}$

Summation Operator

The summation operator \sum means we are **adding up a sequence of values**.

Example:

$$\sum_{i=1}^4 i = 1 + 2 + 3 + 4 = 10$$

- The bottom number ($i = 1$) tells us where to start.
- The top number (4) tells us where to stop.
- Each term is added together step by step.

pbinom() and dbinom() as Summations

dbinom(k, n, p) gives the probability of **exactly** k successes:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

pbinom(k, n, p) gives the **cumulative** probability up to k successes:

$$P(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i (1 - p)^{n-i}$$

In words:

- `dbinom(k, n, p)` = single term in binomial formula
- `pbinom(k, n, p)` = sum of `dbinom(i, n, p)` for $i = 0$ to k

$$\text{pbinom}(k, n, p) = \sum_{i=0}^k \text{dbinom}(i, n, p)$$

Using pbinom() in R

To find cumulative probabilities for binomial distribution:

- $P(X \leq k)$: `pbinom(k, size = n, prob = p)`
- $P(X > k)$: `1 - pbinom(k, size = n, prob = p)`
- $P(X < k)$: `pbinom(k-1, size = n, prob = p)`
- $P(X \geq k)$: `1 - pbinom(k-1, size = n, prob = p)`

For exact probabilities:

- $P(X = k)$: `dbinom(k, size = n, prob = p)`

Binomial Distribution Examples in R

Example: Coin flipping

- What's the probability of getting 7 or fewer heads in 10 flips of a fair coin?

```
# P(X <= 7) where X ~ Binomial(10, 0.5)
pbinom(7, size = 10, prob = 0.5)
# Output: 0.9453
```

```
# P(X = 5) exactly 5 heads
dbinom(5, size = 10, prob = 0.5)
# Output: 0.2461
```

More Binomial Examples

Example: Quality control

- In a batch of 20 items with 5% defect rate, what's the probability of finding 2 or fewer defective items?

```
# P(X <= 2) where X ~ Binomial(20, 0.05)
pbinom(2, size = 20, prob = 0.05)
# Output: 0.9245
```

```
# P(X > 2) = 1 - P(X <= 2)
1 - pbinom(2, size = 20, prob = 0.05)
# Output: 0.0755
```

Questions?