Statistical distributions

AQA A-level Mathematics 7357, section N: discrete distributions, the binomial and normal distributions

Two models to know

	Binomial $B(n, p)$	**Normal** $N(\mu, \sigma^2)$
Values	$0, 1, \ldots, n$ (a count)	any real number
Probability	$\binom{n}{k} p^k (1-p)^{n-k}$	area under the curve
Mean	np	μ
Variance	$np(1-p)$	σ^2

A discrete probability distribution

x	**1**	**2**	**3**
$P(X = x) = kx$	k	$2k$	$3k$
with $k = \frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

$6k = 1$ so $k = \frac{1}{6}$; $E(X) = \frac{1}{6} + \frac{4}{6} + \frac{9}{6} = \frac{14}{6} = \frac{7}{3}$.

When is a count binomial?

Binomial

A fixed number of **trials**, n.

Each trial a success or a failure.

Trials independent, with constant p.

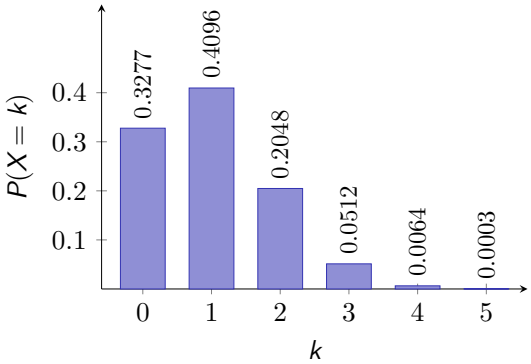
Not binomial

Drawing without replacement from a small bag: p changes.

Throwing until the first six: no fixed n.

Faults that cluster together: not independent.

$B(5, 0.2)$, drawn



- $P(X = 0) = 0.8^5 = 0.328.$
- $E(X) = np = 1,$
  $\mathrm{Var}(X) = np(1-p) = 0.8.$
- Most likely value: 1.

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

Worked: binomial probabilities

1. **Given** $X \sim B(5, 0.2)$
2. **Exactly 1** $\binom{5}{1}(0.2)(0.8)^4 = 5 \times 0.2 \times 0.4096 = 0.4096$
3. **At most 1** $P(X \leq 1) = 0.3277 + 0.4096 = 0.7373$
4. **At least 2** $P(X \geq 2) = 1 - P(X \leq 1) = 0.2627$

Decision

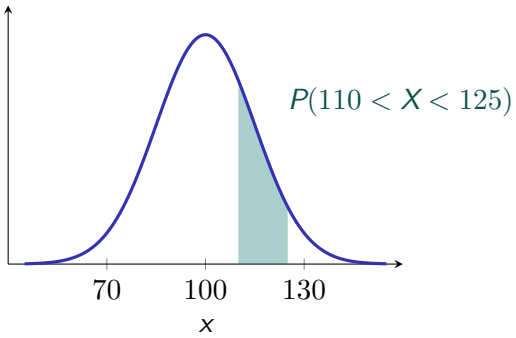
Use the calculator's binomial cdf for sums of many terms.

Translating the words

Words	**Inequality**	**Using** $P(X \leq k)$
at most 4	$X \leq 4$	$P(X \leq 4)$
fewer than 4	$X \leq 3$	$P(X \leq 3)$
at least 4	$X \geq 4$	$1 - P(X \leq 3)$
more than 4	$X \geq 5$	$1 - P(X \leq 4)$

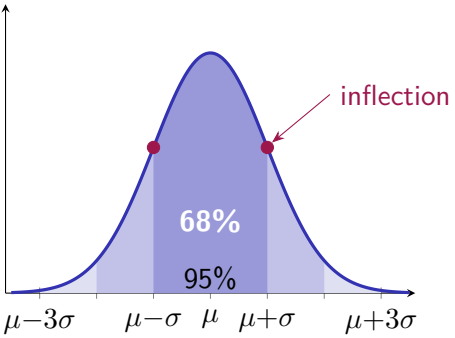
$X \sim B(20, 0.3)$: $P(X < 3) = P(X \leq 2) = 0.0355$.

Continuous variables live in areas



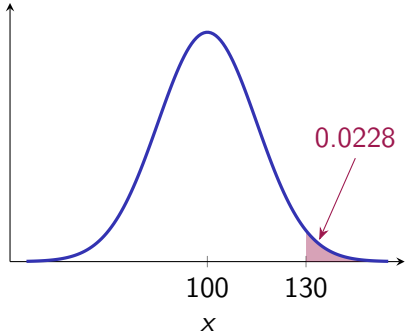
- Probability is area under a density curve.
- $P(X = 130) = 0$ for any exact value.
- So $<$ and $\leq$ give the same answer.

The normal curve



- Symmetric about μ; total area 1.
- Points of inflection at $\mu \pm \sigma$.
- About 68%, 95%, 99.7% within 1, 2, 3 SDs.

Standardise to find a tail



- $X \sim N(100, 15^2)$; find $P(X > 130)$.
- $z = \frac{130-100}{15} = 2$.
- $P(Z > 2) = 0.0228$.

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Probabilities of intervals

1. **Given** $X \sim N(100, 15^2)$
2. **Within 1 SD** $P(85 < X < 115) = 0.6827$
3. **Below 110** $P(X < 110) = P(Z < 0.667) = 0.7475$
4. **Above 85** $P(X > 85) = P(Z > -1) = 0.8413$, by symmetry

Decision

Each answer is a shaded area: sketch it first.

The inverse normal

1. **Given** $X \sim N(100, 15^2)$; find the top 10% boundary
2. **Area to the left** $P(X < x) = 0.9$
3. **Standard value** $z = 1.2816$
4. **Back to x** $x = \mu + z\sigma = 100 + 1.2816 \times 15 = 119.2$

Decision

A score of about 119 puts someone in the top 10%.

$$x = \mu + z\sigma$$

Finding an unknown mean

1. **Given** $X \sim N(\mu, 4^2)$, $P(X < 50) = 0.2$
2. **z value** $P(Z < z) = 0.2$: $z = -0.8416$
3. **Equation** $\frac{50-\mu}{4} = -0.8416$
4. **Solve** $\mu = 50 + 4 \times 0.8416 = 53.37$

Decision

The negative z puts 50 below the mean, as a 20% left tail must.

Choosing the model

Situation	**Model**
Number of faulty items in a batch of 50, each 3% faulty	$B(50, 0.03)$
Masses of apples from one orchard	normal
Number of sixes in 30 throws	$B(30, \frac{1}{6})$
Drawing 3 cards from 10 without replacement	neither: p changes

For large n, $B(n, p)$ is close to $N(np, np(1-p))$.

Check yourself

Question	**Answer**
Mean and variance of $B(40, 0.25)$?	10 and 7.5
"More than 6" as a cdf?	$1 - P(X \leq 6)$
$P(X = 12)$ for a normal X?	0
z for $x = 88$ in $N(100, 15^2)$?	-0.8
Fraction within 2σ of μ?	about 95%