<u>Lecture 3</u>: Probability and Statistics

Let us continue our review of the "classical" theory of probability and statistics in order to firm up the foundations, and increase our understanding of the contrast with "quantum probability amplitudes"

Discrete vs. continuous probability distributions

Let the set of possible outcomes be

<u>countable</u> and finite $\quad S = \{ x_i \mid i = 1, \ldots, N \}$

Let $\quad p_i = P(x_i) \geq 0$ (probability of i^{th} outcome)

Normalization $\quad \sum_{i=1}^{N} p_i = 1$

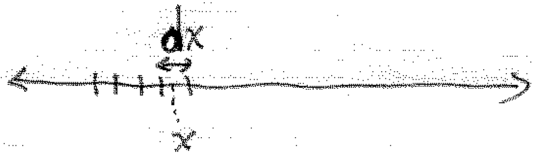
Note: If $\sum_i p_i > 1$, we can always normalize the probability distribution by defining

$$\tilde{p}_i = \frac{p_i}{\sum_i^N p_i} \quad \Rightarrow \quad \sum_i \tilde{p}_i = 1$$

This can all be extended to countably <u>infinite</u> sets $\quad N \to \infty \quad \sum_{i=1}^{\infty} p_i \to 1$ (infinite series converges)

<u>Continuous variables</u>: Suppose now that the set S has an uncountable # of outcomes, e.g. position along a line x, $-\infty \leq x \leq \infty$ (real number)

To "discretize" the outcomes, we break the real line into countable intervals Δx and then take the limit as $\Delta x \to dx$ (differential)



Define $p(x)\,dx$ = probability to have value in range dx around x

$\Rightarrow$ $p(x)$ is a probability density

Units $[p(x)] = \frac{1}{[x]}$ eg. position $[p(x)] = \frac{1}{\text{Length}}$

$\Rightarrow$ By rule $p(A \cup B) = p(A) + p(B)$

The probability to be in a finite interval

$$P(x \in [a,b]) = \int_a^b dx\, p(x)$$

If the sample space is the entire real line

Normalization $\Rightarrow \int_{-\infty}^{\infty} dx\, p(x) = \underline{1}$

Generally, we can go between discrete and continuous prob. distributions through the substitution $\sum_i \rightarrow \int dx$

Joint probability distributions

$p(A, B) \equiv p(A \text{ and } B)$

If A and B are countable sets $A = \{a_i\}$, $B = \{b_j\}$

$$\Rightarrow \text{Normalization} \quad \sum_{i=1}^{i_{max}} \sum_{j=1}^{j_{max}} p(a_i, b_j) = 1$$

If A and B are continuous variables; e.g
A = position on x-axis, B = position on y-axis

$\Rightarrow$ Joint probability density $p(x,y)$ $\left[\text{units } \frac{1}{\text{Area}}\right]$

$p(x,y)\,dx\,dy$ = probability for outcome (x,y) in patch of area $dx\,dy$

Normalization: $\int dx\,dy\; p(x,y) = 1$

Conditional distributions

$$p(a,b) = p(a|b)\,p(b) = p(b|a)\,p(a)$$

If statistically independent

$$\Rightarrow p(a|b) = p(a) \qquad p(b|a) = p(b)$$

$$\Rightarrow p(a,b) = p(a)\,p(b)$$

Marginal distributions

Given a joint distrubution (say continuous)

$$p(x,y)$$

$$p(x) \equiv \int dy\; p(x,y) \qquad p(y) \equiv \int dx\; p(x,y)$$

$$= \int dy\; p(x|y)\,p(y) \qquad = \int dx\; p(y|x)\,p(x)$$

Bayes' theorem

If probabilities represent our best guess given prior knowledge, than we need an update rule once we gain new information. This comes in the form of Bayes theorem.

Recall $p(x,y) = p(x|y)\,p(y) = p(y|x)\,p(x)$

$$\Rightarrow p(x|y) = \frac{p(y|x)\,p(x)}{p(y)} = \frac{p(y|x)\,p(x)}{\sum_x p(x,y)}$$

$$\boxed{p(x|y) = \frac{p(y|x)\,p(x)}{\sum_x p(y|x)\,p(x)}}$$

In words: Suppose we have ~~same~~ prior probability distribution for x $p(x)$ and we learn y is true, what is $p(x|y)$?

Bayes theorem says multiply the prior $p(x)$ by the "likelyhood" that y is true given x and then renormalize.

The notion of updating your probability assignment based on new information is crucial for understanding quantum measurement

Expected value

The mathematical expectation value is your best guess given the current information represented by a probability distribution.

Given set $\{x_i\}$

$$\langle x \rangle = \begin{cases} \sum_i x_i \, p(x_i) & \text{discrete} \\ \int dx \; x \, p(x) & \text{continuous} \end{cases}$$

Note, for discrete finite set with N element, when all probabilities are equal $p(x_i) = \frac{1}{N}$

$$\Rightarrow \quad \langle x \rangle = \frac{1}{N} \sum_i x_i \quad \text{(arithmetic mean)}$$

- Functions of a random variable : $f(x)$

$$\langle f(x) \rangle = \sum_i x_i f(x_i) = \int dx \; x f(x)$$

Linearity of expectation value

$$f(x) = c_1 f_1(x) + c_2 f_2(x)$$

constants

$$\Rightarrow \langle f(x) \rangle = c_1 \langle f_1(x) \rangle + c_2 \langle f_2(x) \rangle$$

- Expected values for joint prob. distributions

Given $p(X, Y)$ $\quad \langle f(X,Y) \rangle = \sum_{i,j} p(x_i, y_j) f(x_i, y_j)$

$$= \int dx dy \; p(x,y) f(x,y)$$

Uncertainty

How certain are we of the expected value? One way to quantify this is to look at the expected deviation of an outcome from the expected value. Since we don't care (necessarily) about the sign of the deviation, it is natural to consider:

Variance: $\sigma^2 \equiv \langle (x - \langle x \rangle)^2 \rangle$

$$\sigma_x^2 = \langle x^2 - 2x \underbrace{\langle x \rangle}_{\text{constant numbers}} + \underbrace{\langle x \rangle^2}\rangle = \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2$$

$$\Rightarrow \boxed{\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2}$$

The uncertainty is the square root of the variance (sometimes known as the rms for root mean square or standard deviation)

$$\sigma_x \equiv \sqrt{\langle (x - \langle x \rangle)^2 \rangle} = \sqrt{\langle x^2 \rangle - \langle x \rangle}$$

Correlations

If two events are unrelated we say that are uncorrelated; their joint probability distribution factors for statistically independent events:

$$p(x,y) = p(x)\,p(y)$$

$\Downarrow$ Iff

$$\Rightarrow \langle XY \rangle = \langle X \rangle \langle Y \rangle$$

Generally: Covariance (correlation) $\sigma_{xy}^2 = \langle XY \rangle - \langle X \rangle \langle Y \rangle$

Examples

Gaussian:

One of the most important (and ubiquitous) continuous probability distrubition is the so called "Gaussian" or "normal" distribution.

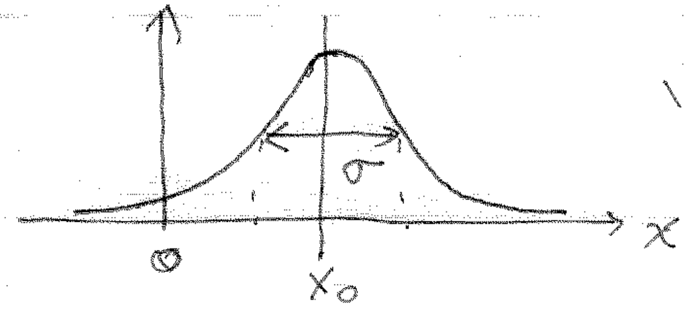
Normalized $p(x) = \frac{1}{\sqrt{2\pi(\Delta x)^2}} e^{-\frac{(x-x_0)^2}{2(\Delta x)^2}}$

Direct integration shows:

$$\langle x \rangle = \int_{-\infty}^{\infty} dx \; x \; p(x) = x_0$$

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \int_{-\infty}^{\infty} dx \; x^2 p(x) - x_0^2$$
$$= (\Delta x)^2$$

Thus, the Gaussian is parameterized by just to things, the mean and variance of the distribution



"Bell curve"

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\langle x \rangle)^2/2\sigma^2}$$

Examples:

- Discrete distribution: "Bernoulli trial"

  Coin flip - Two values of x $\begin{cases} \text{heads} = 1 \\ \text{tails} = 0 \end{cases}$

  Let $p = P(x=1)$ $\quad q = P(x=0)$

  $p + q = 1$

  $\langle x \rangle = p$, $\quad \langle x^2 \rangle = p \Rightarrow \sigma^2 = p - p^2$
  $= p(1-p) = pq$

- Binomial

  Suppose we have N coins that are all weighted the same way and thrown in the same "random" way (identical random variables)

Let X be the number of heads $\quad X = \sum_{i=1}^{N} x_i$

We seek the probability that $X = M$, (i.e. M heads). For a given sequence to come up with M heads the probability is $p^M q^{N-M}$

(To see this, consider example where the first M coins are heads and the last $N-M$ tails.

$$p(x_1=1, x_2=1, \ldots x_M = 1, x_{M+1}=0, x_{M+2}=0, \ldots, x_N = 0)$$

$$= \prod_{i=1}^{M} p(1) \prod_{i=M+1}^{N-M} p(0) = p^M q^{N-M}$$

(Next Page)

Now, there are $\binom{N}{M} = \frac{N!}{M!\,(N-M)!}$ different sequences which have M heads, and all are statistically independent

$$\Longrightarrow \boxed{P_{\text{binomial}}(X = M) = \binom{N}{M} p^N \underbrace{(1-p)}_{=q}{}^{N-M}}$$

This is known as the binomial distribution. It is clearly normalized since

$$\sum_{M=0}^{N} P(X=M) = \sum_{M=0}^{N} \binom{N}{M} p^N q^{N-m} = (p+q)^N$$

"Binomial expansion"

$$= (1)^N = 1 \checkmark$$

An important physics application is the random walk, with application to Brownian motion and diffusion.

Let a particle move "randomly" along a line with step of length l either to the right or left. Let p = ~~probabl~~ probability to move right in a step, $q = 1-p$ = prob. to move left.

Then, after N steps, the probability to have moved M steps to the right is $P_{\text{binomial}}(M)$.

Expected value:

$$\langle m\rangle = \sum_{m=0}^{N} m\,P(m) = \sum_{m=0}^{N} m\binom{N}{m}p^m q^{N-m} = \left(\sum_{m=0}^{N}\binom{N}{m}x^m\right)q^N$$

$q = 1-p$ where $x = p/q$

Aside:

Trick: $(1+x)^N = \sum_{m=0}^{\infty}\binom{N}{m}x^m$ (binomial expansion)

$$\Rightarrow \frac{d}{dx}(1+x)^N = N(1+x)^{N-1} = \sum_{m=0}^{N}\binom{N}{m} m\, x^{m-1}$$

$$\Rightarrow \sum_{m=0}^{N}\binom{N}{m} m\, x^m = N x(1+x)^{N-1} = N\frac{p}{q}\left(1+\frac{p}{q}\right)^{N-1}$$

$$\Rightarrow \langle m\rangle = Np\left(\frac{q+p}{q}\right)^{N-1} q^{N-1} = Np\,(q+p)^N$$

$$\Rightarrow \boxed{\langle m\rangle = Np}$$ Whew!

(there's actually a much easier way to calculate this based on momentum generating functions)

Similarly: $\langle m^2\rangle = \sum_{m=0}^{N} m^2 P(m) = N(N-1)p^2 + Np$

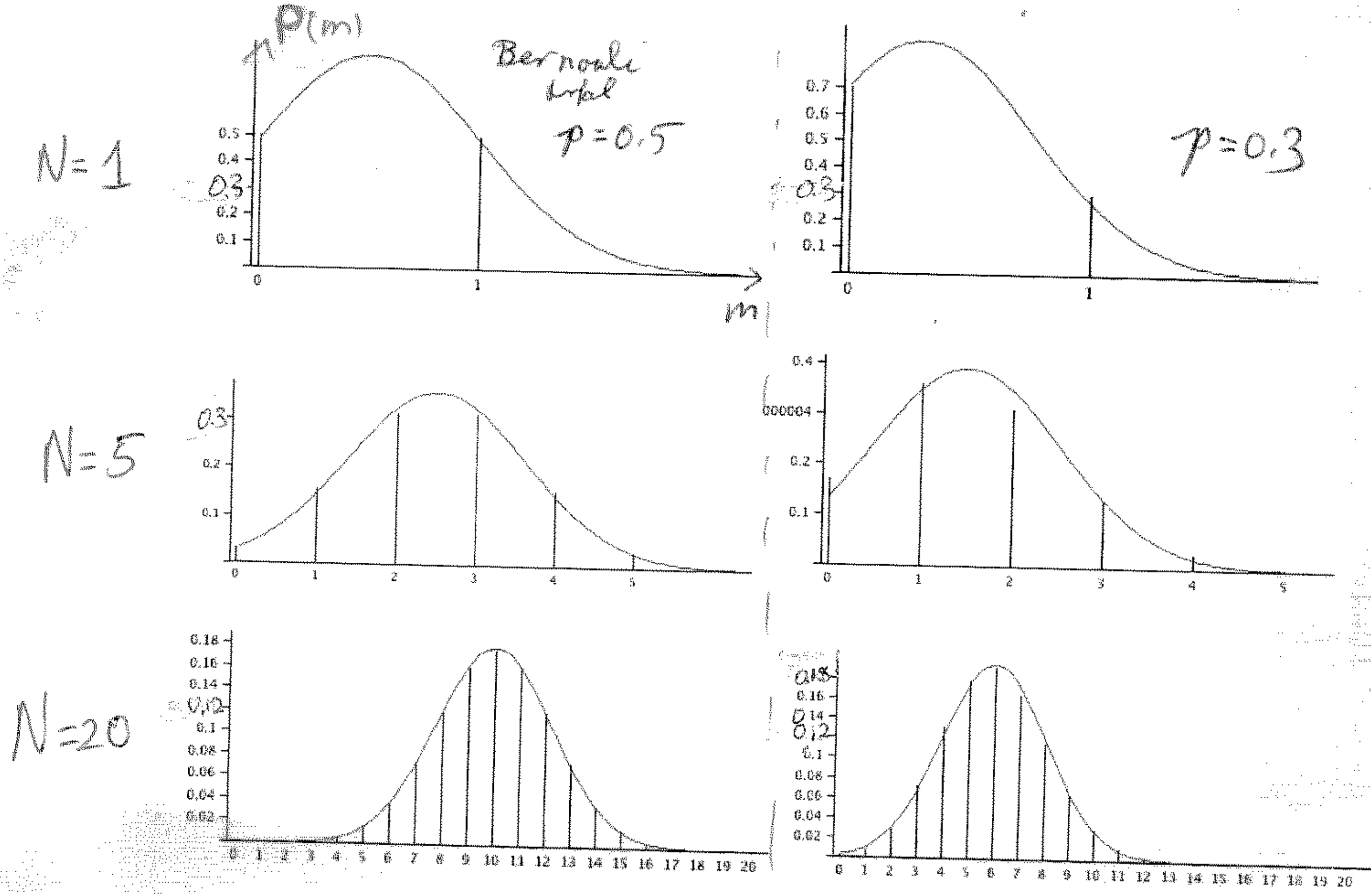
$$\Rightarrow \sigma^2 = \langle m^2\rangle - \langle m\rangle^2 = N(p-p^2) = Np(1-p)$$

$$\Rightarrow \boxed{\sigma^2 = Npq}$$

Note: These results are a special case of the general result: Given N stat. independent identical random variables $\{x_i \mid i = 1\cdots N\}$ Let $X = \sum_{i=1}^{N} x_i$

$$\Rightarrow \langle X\rangle = N\langle x\rangle \qquad \sigma_X^2 = N\sigma_x^2$$

Sketch: Binomial distributions



Note: as N increases the distribution becomes more and more <u>Gaussian</u>. This is an example of the <u>central limit theorem</u>. Moreover as $N \to \infty$ the distribution becomes more and more peaked, approaching a "delta function" (more on this soon). Thus as $N \to \infty$

$$(m = Np) \to 1 \qquad P(m \neq Np) \to 0$$

$$\to \quad \boxed{\frac{m}{N} \to p \quad \text{as } N \to \infty}$$

"Empirical probability". N repeated trials. Fraction gives probability