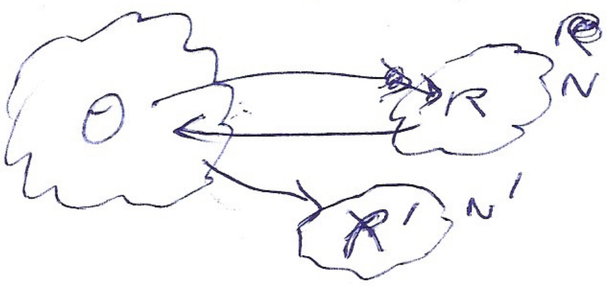
Analysis of problem:

Suppose you pulled a slot machine 5 times, and won 1 time. What's the ~~estimate~~ probability for next pull being a win?

Provides insight into:

for flat prior, MAP and ML estimates are $\frac{1}{5}$. But if we consider full posterior, you ~~set~~ ~~a more lot~~ detailed could do more ~~things~~ like taking average of ~~posterior~~ values based on posterior distribution



Facts:

Beta function $B(x,y) = \dfrac{(x-1)!\,(y-1)!}{(x+y-1)!}$

Beta dist $\pi(\theta; x, y) = \dfrac{1}{B(x,y)} \theta^{x-1}(1-\theta)^{y-1}$

useful Integral: $\displaystyle\int_{\theta=0}^{1} \theta^{r}(1-\theta)^{n-r}\, d\theta = \frac{r!\,(n-r)!}{(n+1)!}$

Slot machine has bernoulli model for one try and hence binomial for n trials. ~~# r are successes, then~~

Let r be # of successes.

Let's use Beta dist $\pi(\theta; \alpha, \beta)$ as prior.

$f_\theta(\theta) = \pi(\theta; \alpha, \beta)$

$P(X|\theta) = {}^{n}C_{r}\, \theta^{r}(1-\theta)^{n-r}$

To understand this consider θ taking a discrete set of values with some probability distribution

θ	$P_\theta(\theta)$
0.2	0.1
0.5	0.2
0.7	0.7
0.9	0.0

② JPT for bernoulli

* $x = 1$ win
  $= 0$ lose

θ

θ	$p(\theta)$
0.2	(0.1)

$x \backslash \theta$	0.2	0.5	0.7	0.9
0	1-0.2	1-0.5	1-0.7	1-0.9
1				

$x \backslash \theta$	0.2	0.5	0.7	0.9
0	1-0.4	1-0.2	1-0.7	
1	0.1	0.2	0.7	

$x \backslash \theta$	0.2	0.5	0.7	0.9
			$(1-0.7)(0.7)$	$(1-0.9)(0)$
0	$(1-0.2)(0.1)$	$(1-0.5)(0.2)$	$(0.7)(0.7)$	0.9×0
1	$(0.2)(0.1)$	$(0.5)(0.2)$		$0 = 1$
Total	0.1	0.2	0.7	

So for one attempt, success is (3pg) | failure is $P(r=0|\theta) = (1-\theta)\, P(\theta)$

$\cancel{P(\theta}$ $\quad f_{r|\theta}(r=1|\theta) = \theta\, f_\theta(\theta)$ $\qquad f_{r|\theta}$

For r ~~attempts~~

$$f_{R|\theta}(r|\theta) = \theta^r (1-\theta)^{n-r}\, \cancel{p_\theta(\theta)}$$

So, likelihood for r wins

$$f_{r|\theta}(r|\theta) = {}^n C_r\, \theta^r (1-\theta)^{n-r}\, \cancel{f_\theta(\theta)}$$

$$f_R(r) = \cancel{{}^nC_r \frac{1}{B(\alpha,\beta)}} \int_\theta f_{r|\theta}(r|\theta)\, f_\theta(\theta)\, d\theta$$

$$= \int_\theta {}^nC_r\, \theta^r (1-\theta)^{n-r} \frac{1}{B(\alpha,\beta)}\, \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$= {}^nC_r \frac{1}{B(\alpha,\beta)} \int_{\theta=0}^{1} \theta^{r+\alpha-1} (1-\theta)^{n-r+\beta-1}\, d\theta$$

$$= {}^nC_r \frac{1}{B(\alpha,\beta)} \frac{(r+\alpha-1)!\,(n-r+\beta-1)!}{(n+\alpha+\beta-1)!}$$

(4)

$$f_{\theta|R}(\theta|r) = \frac{f_{R|\theta}(r|\theta)\, f_\theta(\theta)}{f_R(r)}$$

$$= \frac{{}^nC_r\, \theta^r (1-\theta)^{n-r} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}}{{}^nC_r \frac{1}{B(\alpha,\beta)} \frac{(r+\alpha-1)!\,(n-r+\beta-1)!}{(n+\alpha+\beta-1)!}}$$

$$= \theta^{r+\alpha-1}(1-\theta)^{n-r+\beta-1} \frac{(n+\alpha+\beta-1)!}{(r+\alpha-1)!\,(n-r+\beta-1)!}$$

$$= \frac{\theta^{x-1}(1-\theta)^{y-1}}{\frac{(x-1)!\,(y-1)!}{(x+y-1)!}} \qquad \cancel{\frac{(x+y-1)!}{(x-1)!}}$$

where $x = r+\alpha$
$y = n-r+\beta$

$$= \pi(r+\alpha,\ n-r+\beta)$$

Pg 5

$$p(\overset{\gamma_{n+1}}{\tilde{x}} \mid R) = \int_\theta f_{R,\theta}(\tilde{x}, \theta \mid x)\, d\theta$$

$$= \int f_{R|\theta}(\tilde{x} \mid \theta, R)\, f_{\theta|R}(\theta \mid R)\, d\theta$$

$$= \int f_{R|\theta}(\tilde{x} \mid \theta)\, f_{\theta|R}(\theta \mid R)\, d\theta$$

(1) Markov assumption

$$= \int \theta \; \pi(\gamma+\alpha, n-\gamma+\beta)\, d\theta$$

$$= E[\theta \mid R] = \frac{\gamma+\alpha}{\gamma+\alpha+n-\gamma+\beta}$$

$$= \frac{\gamma+\alpha}{n+\alpha+\beta}$$

For uniform prior $\alpha=\beta=1$

$$p(\tilde{x} \mid R) = \frac{\gamma+1}{n+2} = \frac{1+1}{5+2} = \frac{2}{7}$$

Eqn (1) can be used other ways too:

For ML estimate

$$f_{\tilde{r}|R}(\tilde{r}|R) = \int_\theta f_{R|\theta}(\tilde{r}|\theta)\, f_{\theta|R}(\theta|r)\, d\theta$$

$$\approx \int_\theta f_{R|\theta}(\tilde{r}|\hat{\theta}_{ML})\, f_{\theta|R}(\theta|r)\, d\theta$$

$$= f_{\tilde{r}|\theta}(\tilde{r}|\hat{\theta}_{ML}) \underbrace{\int_\theta f_{\theta|R}(\theta|r)\, d\theta}_{=1}$$

Or for MAP, the same

$$f_{\tilde{r}|R}(\tilde{r}|R) \approx f_{R|\theta}(\tilde{r}|\hat{\theta}_{MAP})$$

Note: In Eqn (1) we said markov assumption (which is essentially markov property). But since all data is conditionally independent given θ

$P(x|\theta, D_n) = P(x|\theta)$

1.

Version 2

First some facts:

Beta function $\qquad B(x,y) = \dfrac{(x-1)!\,(y-1)!}{(x+y-1)!}$

Beta distribution $\qquad \pi(\theta; x, y) = \dfrac{1}{B(x,y)} \theta^{x-1} (1-\theta)^{y-1}$

Useful Integral $\qquad \int_{\theta=0}^{1} \theta^{r} (1-\theta)^{n-r} d\theta = \dfrac{r!\,(n-r)!}{(n+1)!}$

Model:

A single trial of slot machine is a bernoulli trial with θ = probability of success/winning so n pulls/trials with r successes/wins is a binomial distribution.

Before pulling/collecting data of trials, we can assume probability as driven by a beta distribution with α, β chosen to represent our beliefs.

let $R = 1$ single trial success/win
$\quad\;\; = 0$ single trial loss/failure

Let's say we observed n of them

$R_1, R_2 \ldots R_n$. Each is i.i.d

$R_i \sim$ bernoulli(θ)

let R represent any of these.

so

Prior distribution of θ

$$f_\theta(\theta) = \pi(\theta;\ \alpha, \beta) \quad —①$$

Conditional

$$f_{R/\theta}(r/\theta) = {}^nC_r\ \theta^r (1-\theta)^{n-r} \quad —②$$

Here $r = 0$ lose in one trial
$= 1$ win

To understand ② Consider ~~a Sitation~~ a Situation where θ takes discrete values with some probability distribution:

θ	$f_\theta(\theta)$
0.2	0.1
0.5	0.2
0.7	0.7
0.9	0.0

JPT for R, θ

$$P(R=1 \mid \theta=0.2) = \frac{0.2\,\cancel{(0.1)}}{(0.2)\cancel{(0.1)} + (1-0.2)\cancel{(0.1)}} = 0.2 = \theta$$

$R \backslash \theta$	0.2	0.5	0.7	0.9
0	$(1-0.2)(0.1)$ $(1-\theta) f_\theta(\theta)$	$(1-0.5)(0.2)$	$(1-0.7)(0.7)$	$(1-0.9)\,0$
1	$\underline{(0.2)}(0.1)$ $\theta \quad f_\theta(\theta)$	$\underline{(0.5)}\,0.2$ $\theta \quad f_\theta(\theta)$	$\underline{(0.7)}\,(0.7)$ $\theta \quad f_\theta(\theta)$	$\underline{0.9} \times 0$ $\theta \quad f_\theta(\theta)$
Total	0.1	0.2	0.7	0.0

$= 1$

so for one pull, win is characterized

by

$$f_{R|\theta}(r=1 \mid \theta) = \theta$$

$P(R=1 \mid \theta) = \dfrac{\theta\, f(\theta)}{\Sigma\, \theta\, f(\theta)}$

lose is characterized by

$$f_{R|\theta}(r=0 \mid \theta) = (1-\theta)$$

Hence for 'r' wins,

$$f_{R|\theta}(r \mid \theta) = {}^{n}C_{r}\ \theta^{r} (1-\theta)^{n-r}$$

[think of ${}^{n}C_{r}$ addition & probabilities]

(4)

Now

$$f_R(r) = \int_{\theta_1} f_{R,\theta}(r, \theta_1)\, d\theta_1$$

$$= \int_{\theta_1} f_{R|\theta}(r|\theta_1)\, f_{\theta}(\theta_1)\, d\theta_1$$

Integrand is same as

$$= \int_{\theta_1} {}^{n}C_{r}\, \theta_1^{r} (1-\theta_1)^{n-r} \frac{1}{B(\alpha,\beta)} \theta_1^{\alpha-1} (1-\theta_1)^{\beta-1}\, d\theta_1$$

$$= {}^{n}C_{r} \frac{1}{B(\alpha,\beta)} \int_{\theta_1=0}^{1} \theta_1^{r+\alpha-1} (1-\theta_1)^{n-r+\beta-1}\, d\theta_1$$

$$= {}^{n}C_{r} \frac{1}{B(\alpha,\beta)} \frac{(r+\alpha-1)!\,(n-r+\beta-1)!}{(n+\alpha+\beta-1)!} \quad \text{—(4)}$$

Using $\theta = \theta_1$ to illustrate that: all we are doing is integrating across values of θ in its set θ and what variable name we use is irrelevant and not the same as θ used in other places in this document.

Pg 5

$$f_{\theta|R}(\theta|r) = \frac{f_{R|\theta}(r|\theta)\, f_{\theta}(\theta)}{f_R(r)} \quad \text{— (5)}$$

Notice, denominator is

$\int$ numerator $d\theta$ as ~~expect~~ expected

From (1), (2), (4) plugging in (5)

$$f_{\theta|R}(\theta|r) = \frac{{}^nC_r\, \theta^r (1-\theta)^{n-r} \frac{1}{B(\alpha,\beta)}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}}{{}^nC_r \frac{1}{B(\alpha,\beta)} \frac{(r+\alpha-1)!\,(n-r+\beta-1)!}{(n+\alpha+\beta-1)!}} \quad \text{— (6a)}$$

$$= \theta^{r+\alpha-1}(1-\theta)^{n-r+\beta-1} \frac{(n+\alpha+\beta-1)!}{(r+\alpha-1)!\,(n-r+\beta-1)!}$$

$$= \frac{\theta^{x-1}(1-\theta)^{y-1}}{\frac{(x-1)!\,(y-1)!}{(x+y-1)!}}$$ where $x = r+\alpha$, $y = n-r+\beta$

$$= \pi(\theta; x, y) = \pi(\theta; r+\alpha, n-r+\beta) \quad \text{— (6)}$$

To predict next outcome, $\tilde{r} = 1$ win (6)

OR $= 0$ loose

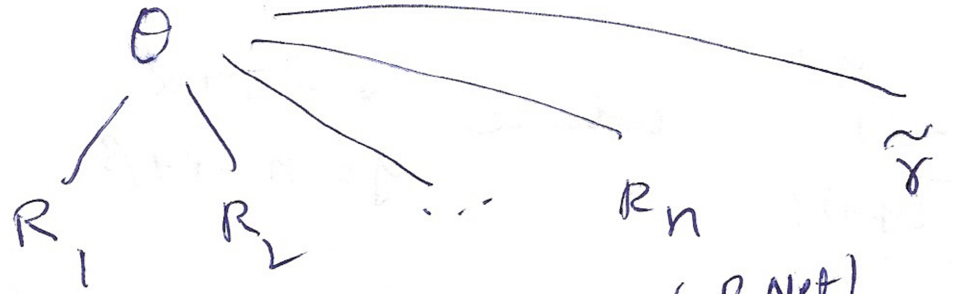
$$f_{\tilde{r}|R}(\tilde{r}|R) = \int_{\theta} f_{R,\theta}(\tilde{r}, \theta | r)\, d\theta$$

$$= \int_{\theta} f_{R|\theta}(\tilde{r} | \theta, \cancel{r})\, f_{\theta|R}(\theta|r)\, d\theta$$

$$= \int_{\theta} f_{R|\theta}(\tilde{r}|\theta)\, f_{\theta|R}(\theta|r)\, d\theta \quad — (7)$$

In (7) we used ~~the~~ fact that each data/outcome is independent of other given θ (markov property)

$$P(\tilde{r} | \theta, R_1, R_2, \dots R_n) = P(\tilde{r}|\theta)$$



(BNet)

$f_{\tilde{\gamma}|R}$ ~~$(\tilde{\gamma}|R) =$~~

Also $\quad f_{R|\theta}(\tilde{\gamma}|\theta) = \theta^{\tilde{\gamma}}(1-\theta)^{1-\tilde{\gamma}}$ — 8

Here since we are considering next pull only, $\tilde{\gamma} = 1$

$$f_{R|\theta}(\tilde{\gamma}=1|\theta) = \theta \quad \text{— 8a}$$

this is nothing but expectation $\theta|R$ as $\pi(\ldots)$ is just the prob. dist. of $\theta|R$ see eqn 6

From 7

$$f_{\tilde{\gamma}|R}(\tilde{\gamma}=1|R) = \int_\theta \theta \; \pi(\gamma+\alpha, n-\gamma+\beta)\, d\theta$$

$$= E[\theta|R] = \frac{\gamma+\alpha}{\gamma+\alpha+n-\gamma+\beta}$$

$$= \frac{\gamma+\alpha}{n+\alpha+\beta} \quad \text{— 9}$$

For uniform prior on θ, $\alpha = \beta = 1$

$$f_{\tilde{\gamma}|R}(\tilde{\gamma}=1|R) = \frac{\gamma+1}{n+2} \quad \text{— 10}$$

Since

$n = 5$

$r = 1$

The probability of next pull being winner, given in the last 5 pulls we got only ~~a~~ one win is

$$= \frac{1+1}{5+2} = \frac{2}{7}$$ plugging in Eqn (10)

Incidentally equation (7) can be used

Other ways:

For max. likelihood

$$f_{\tilde{r}|R}(\tilde{r}|R) = \int_\theta f_{R|\theta}(\tilde{r}|\theta)\, f_{\theta|R}(\theta|r)\, d\theta$$

$$\approx \int_\theta f_{R|\theta}(\tilde{r}\,|\,\hat{\theta}_{ML})\, f_{\theta|R}(\theta|r)\, d\theta$$

$$= f_{R|\theta}(\tilde{r}\,|\,\hat{\theta}_{ML}) \underbrace{\int_\theta f_{\theta|R}(\theta|r)\, d\theta}_{=1}$$

$$= f_{R|\theta}(\tilde{r}\,|\,\hat{\theta}_{ML})$$

For MAP from 7

$$f_{\tilde{\gamma}|R}(\tilde{\gamma}|R) \simeq f_{R|\theta}(\tilde{\gamma}|\hat{\theta}_{MAP})$$

So, you just take $\hat{\theta}_{ML}$ or $\hat{\theta}_{MAP}$ and plug in 8a

For example for ML,

~~xxx~~ prob. of next pull being winner (ML) $= \hat{\theta}_{ML} = \frac{6}{30} = \frac{1}{5}$

prob. of next pull ~~being~~ being winner (MAP) $= \hat{\theta}_{MAP} = \frac{r}{n} = \frac{1}{5}$

(for flat prior ML = MAP)

contd →

If prior is not flat, MAP is calculated as below (see Holy trinity)

$$L = {}^{n}C_{r}\, \theta^{r} (1-\theta)^{n-r} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

from eqn (6a) needs to maximized

~~Logle and sum~~

$$\log L = \cdots + (r+\alpha-1)\log\theta + (n-r+\beta-1)\log(1-\theta)$$

$$\frac{\partial \log L}{\partial\theta} = 0 \Rightarrow \frac{r+\alpha-1}{\theta} = \frac{n-r+\beta-1}{1-\theta}$$

or $\frac{1-\theta}{\theta} = \frac{n-r+\beta-1}{r+\alpha-1}$

or $\hat{\theta}_{MAP} = \frac{r+\alpha-1}{n+\alpha+\beta-2}$ $\left[= \frac{r}{n} \text{ for flat prior } \alpha=\beta=1 \right]$

What about the expected payoff of next slot machine pull? i.e,

~~$E[R]$ $E[R]$ $E[$~~

$$E[\tilde{R} \mid R] = ? \qquad \because R = 0 \text{ or } 1$$

$$E[\tilde{R} \mid R] = 1 \cdot f_{\tilde{R}|R}(\tilde{r} = 1 \mid R) + 0 \cdot f_{\tilde{R}|R}(\tilde{r} = 0 \mid R) \quad (\text{using } \tilde{R} = \tilde{r})$$

$$= \frac{r+1}{n+2}$$

$$= \frac{2}{7}$$