# Connjugate Prior

Why are they useful and proofs of thier righteousness

Ha Vu

07/26/2023

# Contents

1	Why conjugate prior is nice?	2	
	1.1 What is conjugate prior in the first place?	2	
	1.2 Why is it nice?	2	
	1.3 Examples of conjugate prior cases	2	
2	Bionomial likelihood with Beta prior	3	
3	Poisson likelihood with Gamma prior	4	
4 Two cases for the Gaussian distribution, aka more reason why Gaussian runs the world			
5	Multinomial likelihood with Dirichlet prior	12	
6	Exponential likelihood with Gamma prior	13	
7	Geometric likelihood with Beta prior	14	

### 1 Why conjugate prior is nice?

#### 1.1 What is conjugate prior in the first place?

In Bayesian inference, X represents the observed variable, which has a distribution that is parameterized by  $\theta$ , i.e.  $X \sim f_{\theta}(x)$ . Bayesian inference assumes that  $\theta \sim g(\theta)$ , which is called a prior distribution. In most of the cases for application, the goal is to infer the values of  $\theta$  that maximize the posterior probability of  $\theta$ , i.e.  $\arg\max_{\theta} P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int P(X|\hat{\theta})d\hat{\theta}}$  Conjugate prior is the case where the prior distribution of  $P(\theta)$  is chosen so that the posterior distribution  $P(\theta|X)$  is of the same distribution as the prior.

#### 1.2 Why is it nice?

Conjugate is nice because we KNOW analytical form of the posterior distribution.

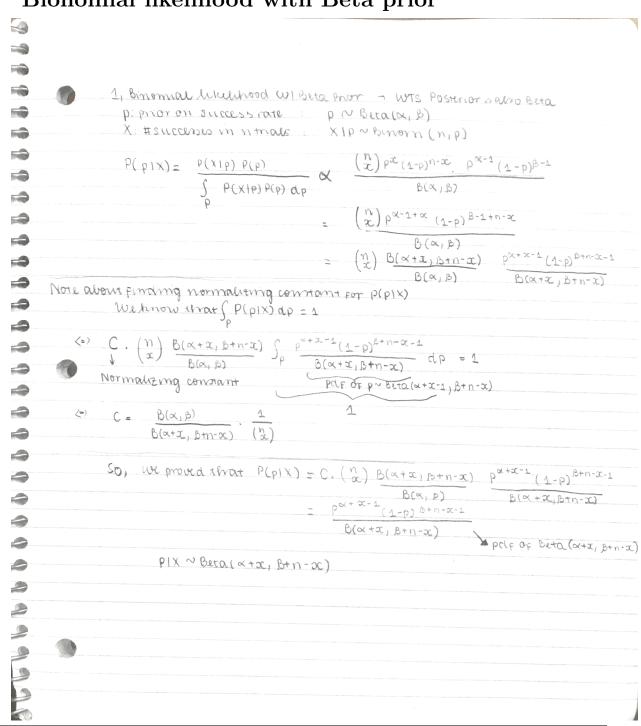
#### 1.3 Examples of conjugate prior cases

I used ChatGPT to list out a list of distribution pairs that will result in a conjugate prior case, and I tried to prove each case.

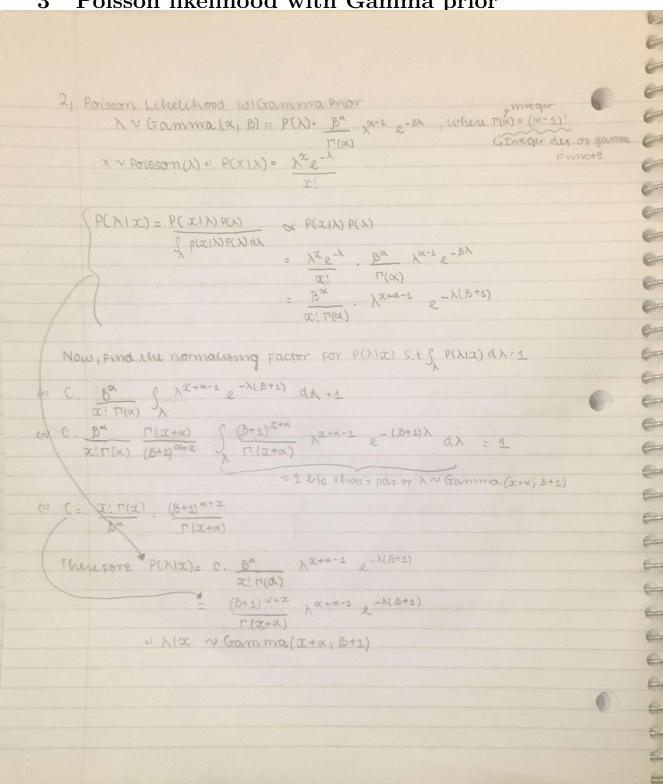
- Binomial likelihood with Beta prior
- Poisson likelihood with Gamma prior
- Gaussian likelihood with Gaussian prior (for the mean  $\mu$ )
- Gaussian likelihood with Inverse Gamma prior (for the variance  $\sigma^2$ )
- Multinomial likelihood with Dirichlet prior
- Exponential likelihood with Gamma prior

• Geometric likelihood with Beta prior

## 2 Bionomial likelihood with Beta prior

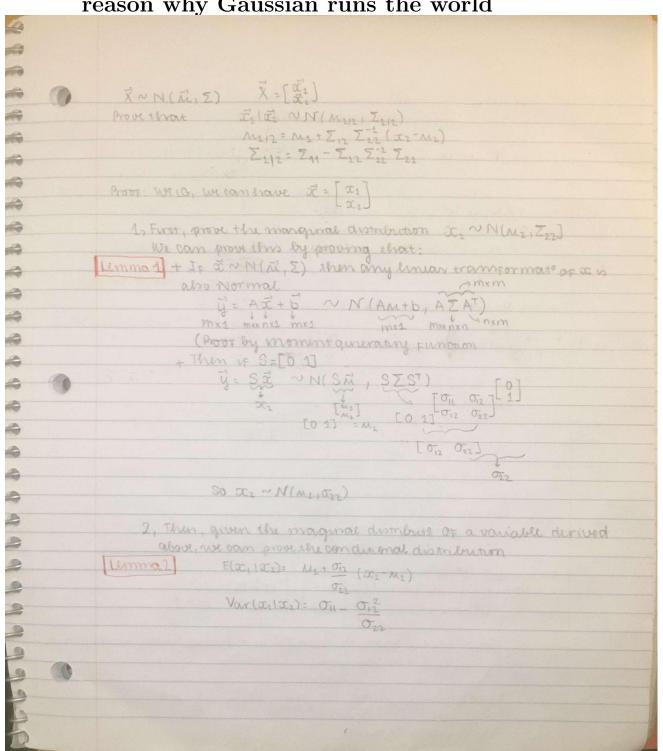


3 Poisson likelihood with Gamma prior



4 Two cases for the Gaussian distribution, aka more

reason why Gaussian runs the world

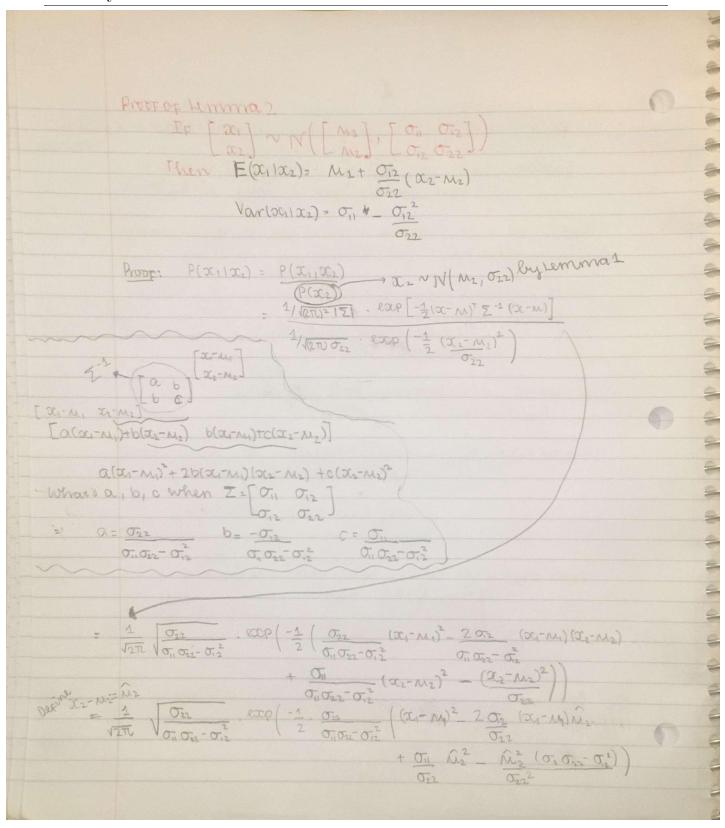


Control of the second	
	3, Now, we can arrive
	m almosa (144 (72)
	X/M Normal (M, O2)
	M ~ Normal (Mo, 002)
	My loan Fr 1, Normal (Efact), Var(x) COULX, MI)
	manual a la manual milking de d
	We have [x] ~ Normal ([E[xt], [Var(x) Cov(x, m)])
	LMJ ([m. ] Ecorlym) Co  + What's E[x]? by previous temma or conditional dist  Weknow SE[xIM] = M = E(x)+ COU(X,M) (M-Mo) (**)  COÚ(xIM) = O=Vor(x) [COU(X,M]² (***)
	LIGHT CETTER TO A FELT COUCKING (1)
	withow telains = ho = elast of (m mo)
	00
	( coύ(α(m)= ο= vor(x) [coυ(x, m]] (**)
	by prer on learning of conditional dist.
	by fire on summin of whom we wise.
	V BOT FERE (VIII) - VOC (C(VIU)) 2
	By low of to tal vanionae: Var[X] = Ey[ Var(X19)] + Vary(E(X19))
	So Var [X] = Em (Var (XIM)) + Varm (E(XIM))
	F (~2) \ \( \( \lambda \)
	EM (O) + Varm (M)
	by low of to tal vanionna: $Var[X] = E_y[Var(X Y)] + Vary(E(X Y))$ So $Var[X] = E_{xx}[Var(X x)] + Var_{xx}(E(X x))$ $= E_{xx}(\sigma^2) + Var_{xx}(x)$ $= Var_{xx}(x) = \sigma^2 + \sigma_0^2 - [Cov(x,x)]^2$ Replace into $(x x)$ $= \sigma^2 + \sigma_0^2 - [Cov(x,x)]^2$
	Replacinto (xx)
	$\sigma^{2} = \sigma^{2} + \sigma_{0}^{2} - [Cov(x, M)]^{2}$
	0 = 0 +00 -[may1yy)
	0°2
	$\langle \Rightarrow \sigma_0^* = [\cos(x, M)]^2$
	O. 2
	(=) $(ov(X, M) = 0.2)$
	Replace COU(X, ru) into (*)
	$M = E(x) + \frac{\sigma_0^2}{\sigma_0^2} (M - M_0)$
-	<i>O</i> <sub>0</sub> <sup>2</sup>
	So [x] ~ Normal [ No], [ 02+0,2 0,2])
	Co [a] a Normal [ [ 1 ] [ 2 2 2 2]
	0 0 0 00 00
	[m] [mo] co oo]
Halle State	
	4 Mars is the apply the miles as conditioned distribution come (2)
	4, Now, if we apply the rules of conditional distribution from (2)
	pla ~ Normal with:
	Elvinia
	- (x-Ma)
	$E[M]xJ = Mo + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} (x - Mo)$
	Var (M/00) = 002 - 004 = 00202
	J2+502 0-8+502
	1,00

9	
O Per Per Per Per Per	
mg.	
meg	5, Can we extend the framework to find P(M/00, a oc)? Yes
	But let's turn P(org oca) M) into P(oc/M)
	$\chi_i \stackrel{\text{Nd}}{\sim} N(n, \sigma^2)$
-	$= \left[ \begin{array}{c} X_1 \\ X_2 \end{array} \right] \sim \mathcal{N}\left( \left[ \begin{array}{c} M \\ M \end{array} \right], \sigma^2 \mathbf{I}_n \right)$
-	LxnJ LMJ
	By the Fact that any linear transformator & normal
	(proved in paint (1))
-	Let S=[1 - 1] then we have:
-	SY NIGHT STOT
-	$S\overrightarrow{X} \sim N(S\overrightarrow{M}, S\overrightarrow{O}T_{0}S^{T})$
-	nxn
-	€ 5" Xi ~ N (n.M, no2)
-	(-) Zinxi ~ N (M, 0)
-	
	Var(ax): a2 Var(x), ais a constant
	$(=)  \times \sim N(M, \sigma^2)$
	Next 10s some set of the set of t
	Next, We want to show that P(x1 - x1/4) & P(x1/4)
	$P(\infty_1 - \alpha_n   n) = TL^n + exp(-1 - (\alpha_i - n)^2)$
000000000000000000000000000000000000000	de exp(-1 7 " (x-21)2)
2	$\alpha_{n} \exp\left(-\frac{1}{2\sigma^{2}}\sum_{i=1}^{n}\left(\alpha_{i}-n_{i}\right)^{2}\right)$
2	$\alpha_{n} \exp\left(-\frac{1}{2\sigma^{2}}\left[\left(\Sigma_{i}^{n}, \alpha_{i}^{2}\right) - 2n\left(\Sigma_{i}^{n}, \alpha_{i}\right) + nn^{2}\right]\right)$
2	
	$ \sqrt{n} \exp\left(\frac{-1}{2\sigma^2}\left(nn^2-2nn.\overline{x}\right)\right) $
-	
9	$\propto_{\mathcal{M}} \exp\left(-\frac{2v}{2\sigma^2}(\overline{x}-\mu)^2\right)$
9	$\alpha_{n} P(\overline{x} n)$
عا	based on sectors) P(MIII In) & P(II) & MIN) P(M) & P(II) P(M) & P(MII)
	=> MIX Normal with
1	$E(M \overline{x}) = M_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma_0^2} (\overline{x} - M_0)  Var(M \overline{x}) = \frac{\sigma_0^2 \sigma_0^2}{\sigma_0^2 + \sigma_0^2}$
(100 COO)	2 + 00

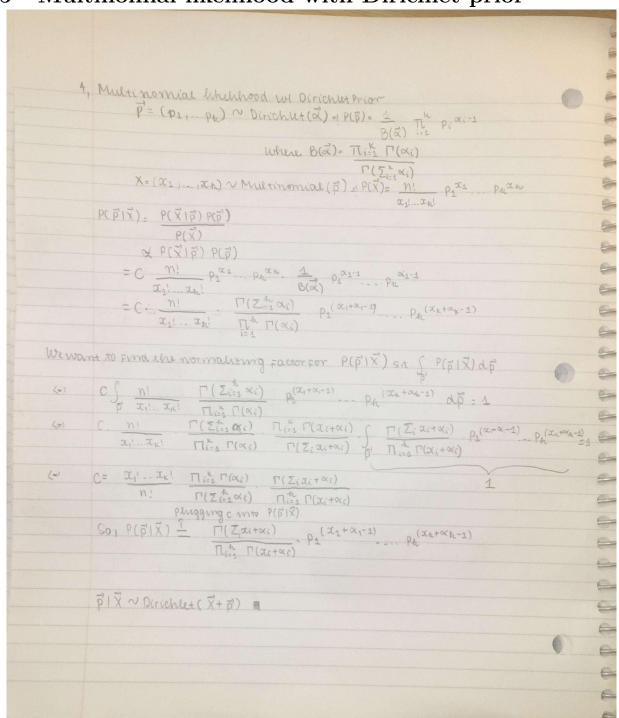
		9
		9
Pom.	se the following statement	
110	T or did her many	
	I = oci W N(n,0°) , where uis fixed	$-1$ $exp(\frac{\beta}{\sigma^2})$
	$\sigma^2 \sim \Gamma(\sigma(\alpha, \beta)) = P(\sigma^2(\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}(\sigma^2)^{\alpha}$	1 exp == 1
	Inverse Gamma (CX)	1013
		9
	Then relation of NTG(x, n, B, 1 51 12)	
	Then 021 x1 x1 ~ IG(x+1/2, B+1/2 \(\int_i(xi-N)^2\)	
		9
	20 (4 5 (4 5 (4 2)	9
	$P(\alpha_1 - \alpha_n   \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{-1}{2\sigma^2} \sum_i  \alpha_i - \mu_i ^2\right)$	
-1	(1270) 20	3
	$P(\sigma^2) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-2} \exp\left(\frac{-\beta}{\sigma^2}\right)$	9
	$\Gamma(\alpha)$ $\sigma^2$	9
	$P(\sigma^2 \mid x_1 \dots x_n) \propto P(x_1 \dots x_n \mid \sigma^2) P(\sigma^2)$	
	$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{-1}{2\sigma^2} \sum_{i} (\alpha_i - M)^2\right) \cdot \frac{B^{\alpha}}{\Gamma(\alpha_i)} (\sigma^2)^{-\alpha-1}$	1 0 m/-B) =
	[202 ] [(X)	14 ( 52 ) =
	0x exp (-1 (12,100 m)2+B)) (02) 2-0-1.	
		0 9
		exp(-B)
	There o2   a2 In ~ IG (x+n , B + 2 Zilai N)2)	0
1-	1 2 2 1	-
		9
		9
		-
		6
		0
		-
4		9
		67
L		

No. of Concession, Name of Street, or other Persons, Name of Street, or ot		
6		
and)	0	Proof of Lemma 1
-		IF X ~ N(12, 2) then any linear transfor mate of X is also
-		Namma
19		Inparticular: $\vec{y} = A\vec{x} + \vec{b} \sim N(A\vec{n} + \vec{b}, A Z A^T)$ mx1 mx1 mx1 mx1 mx1 mx1 mx1 mxn xxn nxm
		mad mannad mad mad man nam nam
1		
-		Proof: We know that the moment generating function of is:
9		$M_{\infty}(t) = E(e^{-1}_{\infty}(t^{T}_{\infty}))$
0		Mack) = Eleaple all
		The a 10 FF. (170 7.5)]
-		Thun Mylt): E[eap(+T(Ax+b))]
-		(17. → (17. → \)7
		$= \exp(t^T b) \cdot E[\exp(t^T A \overline{x})]$
-		$= \exp(t^T \overline{b})  M_{\infty}(A^T t)$
9		
9		The moment generating sunct of \$\vec{x} \nabla N(\vec{x}_0, \vec{x}) is
3	-	$Mx(t) = exp(t^{\dagger}M + \frac{1}{2}t^{\dagger}Zt)$
		2 /
9		Then, we have My(t) = exp(tTb). Mx(ATt)
9		= exp(tTb), exp(tTA Ti+ 1 tAZATt)
9		2
		$= exp(+T(\overline{b}+A\overline{n})+ + +T(A\Sigma AT)+)$
,		My Ey
1		Smoe MGF and PDF are equivalent, we have
-		YN N(B+AM, AZAT)
		J. MOTAM, AZA,
•		
	100	

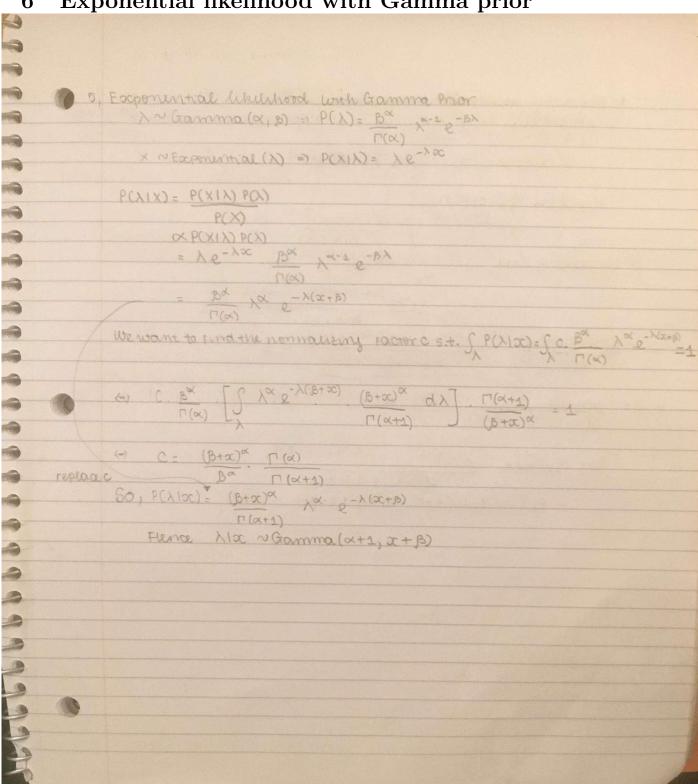


4		72
-		-02:=0
-		sine on 022
-	0	Define $\sigma_{11} \frac{\sigma_{12} - \sigma_{12}}{\sigma_{21}} = \frac{\sigma^{2}}{\sigma_{12}}$ $= \frac{1}{\sqrt{2\pi} \hat{\sigma}} \exp\left(-\frac{1}{2} \cdot \frac{1}{\hat{\sigma}^{2}} \left( \frac{\chi_{1}^{2} - 2}{\sigma_{12}} \cdot \frac{\sigma_{12}}{\sigma_{22}} \cdot \frac{\chi_{1}}{\sigma_{22}} \cdot \chi$
9		12TI ô 2 022 022
-		+ 112 - 2 x1 M1 + OH N2 - M2 (ON O22 - O12)
9		$ \frac{\sqrt{2\pi} \hat{\sigma}}{\sqrt{2\pi}} \left( \frac{2}{\sigma_{22}} \left( \frac{\sigma_{12}}{\sigma_{22}} \frac{\sigma_{12}}{\sigma_{22}} \frac{\sigma_{12}}{\sigma_{22}} \frac{\sigma_{12}}{\sigma_{22}} \frac{\sigma_{12}}{\sigma_{12}} \frac{\sigma_{13}}{\sigma_{12}} \frac{\sigma_{14}}{\sigma_{12}} \frac{\sigma_{12}}{\sigma_{12}} \frac{\sigma_{12}}{\sigma_{12}} \right) \right) $
0		
0		$=\frac{1}{\sqrt{2\pi}\hat{\sigma}}\exp\left(\frac{1}{2\hat{\sigma}^{2}}\left(\frac{\alpha_{1}^{2}}{2\hat{\sigma}^{2}}\left(\frac{\alpha_{1}^{2}}{\sigma_{12}}\frac{\alpha_{1}}{\sigma_{12}}+\frac{\alpha_{1}}{\sigma_{12}}\right)+\frac{\sigma_{1}}{\sigma_{12}}\frac{\alpha_{2}^{2}}{\sigma_{12}}+\frac{2\sigma_{12}}{\sigma_{12}}\frac{\alpha_{1}}{\sigma_{12}}$
9		$M^2$ $M^2$ $(\pi \pi - \sigma^2)$
0		$+ M_1^2 - \hat{M}_2^2 \left( \sigma_1 \sigma_{22} - \sigma_{12}^2 \right) $
0		$-\frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left(\frac{-1}{2\hat{\sigma}^{2}} \left(\frac{\pi_{1}^{2}}{2\pi} - 2\infty_{1} \left(\frac{\sigma_{12}}{\sigma_{12}} + M_{1}\right) + M_{1}^{2} + 2M_{1} \frac{\sigma_{12}}{\sigma_{12}} + M_{1}\right) + M_{1}^{2} + 2M_{1} \frac{\sigma_{12}}{\sigma_{12}} + M_{1}^{2} + M_{1}^{2} + 2M_{1} \frac{\sigma_{12}}{\sigma_{12}} + M_{1}^{2} +$
5		$\sqrt{2\pi}\hat{\sigma}$ $\left(2\hat{\sigma}^2\right)$ $\left(2\hat{\sigma}^2\right)$ $\left(2\hat{\sigma}^2\right)$ $\left(2\hat{\sigma}^2\right)$
9		+ M2 ( OII _ OI O2 - O12)
B		022 022
9		012 012
	0	
;	-	$= \frac{1}{\sqrt{2\pi} \hat{\sigma}} \left( \frac{1}{2\hat{\sigma}^2} \left( \frac{1}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^2} \left( \frac{1}{2\hat{\sigma}^2} + \frac{1}$
		$=\frac{1}{\sqrt{2\pi}\hat{\sigma}}\exp\left(\frac{-1}{2\hat{\sigma}^2}\left[\frac{\chi_2}{\sqrt{2\pi}}\left(\frac{M_1+\sqrt{2\pi}}{\sqrt{2\pi}}\hat{M}_2\right)\right]^2\right).$
		30 x2/x2 ~ N ( N12+ 012 (N2-x2) ) 011 02 - 012 )
	9	

## 5 Multinomial likelihood with Dirichlet prior



Exponential likelihood with Gamma prior 6



Geometric likelihood with Beta prior 7) Geometric Whelihood w/ Beta mor probeta(x, B) => P(p)= px-1 (1-p)B-1 where B(x, B)= T(x)T(B) X; ~ Geometric (p) =1 P(x)= T( (1-p)x0-1 p) P(p|X) & P(X|p)P(p) : (1-p)(Z;00)-K pK px-1(1-p)B-1 we want to find CE.t.  $\int_{P} P(P|X) = 1$   $C. \int_{Q} \frac{(1-p)^{(\Sigma_{i}\alpha_{i})-K+B-1}}{B(\alpha_{i}B)} p^{K+\alpha-1}$ =) C = B(x,B) B(x+K,( [ac)+B-K) So,  $P(p|\vec{X}) = PK+\alpha-2 (1-p)(\Sigma_i\alpha_i) - K+\beta-1$   $B(K+\alpha_1(Z_i\alpha_i)+\beta-K)$   $= P(\vec{X}) \sim Beta(K+\alpha_1(\Sigma_i^{\kappa}\alpha_i) + \beta-K)$